DOCUMENT RESUME

ED 317 577 TM 014 623

| | |
|---|---|
| AUTHOR | Owen, K. |
| TITLE | Bias in Test Items: An Exploration of Item Content and Item Format. Report P-106. |
| INSTITUTION | Human Sciences Research Council, Pretoria (South Africa). |
| REPORT NO | ISBN-0-7969-0771-4 |
| PUB DATE | 89 |
| NOTE | 137p.; An Afrikaans abstract and an English abstract are included. |
| PUB TYPE | Reports - Research/Technical (143) |
| EDRS PRICE | MF01/PC06 Plus Postage. |
| DESCRIPTORS | Afrikaans; Black Students; *Cognitive Tests; Comparative Testing; Content Validity; English; Ethnic Groups; Foreign Countries; Indians; Item Analysis; *Racial Bias; Secondary Education; *Secondary School Students; *Test Bias; Test Format; Testing Problems; Test Items; White Students |
| IDENTIFIERS | *Reasoning Tests; *South Africa |

ABSTRACT

Sources of item bias located in characteristics of the test item were studied in a reasoning test developed in South Africa. Subjects were 1,056 White, 1,063 Indian, and 1,093 Black students from standard 7 in Afrikaans and English schools. Format and content of the 85-item Reasoning Test were manipulated to obtain information about bias or differential group achievement. Bias indices were calculated by the Scheuneman technique and the three-parameter logistical model. The Junior Aptitude Tests, the Raven Standard Progressive Matrices, and a biographical questionnaire were used, along with the Reasoning Test, in this investigation. Results suggest a distinction between true and apparent item bias. True item bias included aspects such as language factors, overly attractive distractors, and a lack of the knowledge and concepts that are usually acquired informally. Apparent bias--a feature of the examinee's test behavior rather than of the item--was indicated as a tendency to use associative, rather than conceptual similarity in classification, selective attention, and an insufficiently logical approach. Findings also suggest that: (1) items in which a statement was made functioned less well than did other types; (2) story-type items led to poor performance; (3) sayings were not suitable open item types; (4) figures were superior to language in terms of content; (5) figural analogies were superior to verbal analogies; (6) series were better than analogies in terms of format; and (7) syllogisms were not suitable item types. Twenty-seven tables and one figure contain study data. (SLD)

***********************************************************************
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
***********************************************************************

ED317577

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. G. GARBERS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

INSTITUTE FOR PSYCHOLOGICAL AND EDUMETRIC RESEARCH P-106

# Bias in test items: An exploration of item content and item format

RGN·HSRC

TM014623

BEST COPY AVAILABLE

**K. Owen**

# Bias in test items: An exploration of item content and item format

Report P-106

# Bias in test items: An exploration of item content and item format

K. Owen

Pretoria
Human Sciences Research Council
1989

K. Owen, D. Litt. et Phil.

Division for Aptitude Research
Institute for Psychological and Edumetric Research
Executive Director: Dr R.J. Prinsloo

ISBN 0 7969 0771 4

© Human Sciences Research Council, 1989

Published by the HSRC
134 Pretorius Street
Pretoria

# ACKNOWLEDGEMENT

Special thanks are due to the following:
the Centre for Language Services of the HSRC which undertook the translation of the original report, Mrs W.G Verster who typed the manuscript and prepared it for publication, and Mrs Jennifer Folwaczny who helped to revise the translation.

CONTENTS

EKSERP . . . . . viii

ABSTRACT . . . . . x

1.0 INTRODUCTION . . . . . 1

2.0 AIM . . . . . 1

3.0 ITEM BIAS . . . . . 2

3.1 Some background factors with regard to item bias . . . . . 2
3.1.1 Culture . . . . . 2
3.1.2 Socio-economic status (SES) . . . . . 4
3.1.3 Language . . . . . 6
3.1.4 Cognitive style . . . . . 7
3.2 Definitions of item bias . . . . . 10
3.3 Biased item types . . . . . 11
3.3.1 Item content . . . . . 11
3.3.2 Item format . . . . . 13
3.4 Hypothesis with regard to item bias . . . . . 14

4.0 RESEARCH DESIGN . . . . . 16

4.1 Samples . . . . . 16
4.2 Measuring instruments . . . . . 16
4.3 Procedures, methods and techniques . . . . . 17

5.0 RESULTS . . . . . 22

5.1 The way in which an item is formulated . . . . . 22
5.1.1 Cues in test items . . . . . 22
5.1.2 Position of the correct answer among the distractors . . 24
5.1.3 Attractiveness of item distractors . . . . . 27
5.1.4 True and false answers . . . . . 29
5.1.5 Positive and negative items . . . . . 31
5.2 Item content . . . . . 33
5.2.1 Concrete and abstract items . . . . . 33
5.2.2 The use of words and letters in test items as opposed to the use of numbers and figures . . . . . 37
5.2.3 Concepts with regard to time . . . . . 41
5.2.4 Direction and orientation . . . . . 44
5.2.5 Mathematical "story" problems and calculations . . . . . 46
5.2.6 The use of statements in test items . . . . . 48
5.2.7 Synonyms and antonyms . . . . . 50
5.3 Item format . . . . . 50
5.4 Information in item distractors . . . . . 54
5.4.1 Illogical answers . . . . . 54
5.4.2 Uncritical attitude . . . . . 54
5.4.3 Ignoring contradictions . . . . . 55
5.4.4 Selective attention . . . . . 55

**6.0 SUMMARY AND DISCUSSION . . . . . . . . . . . . . . . . . . . . 56**

6.1 Formulation of items . . . . . . . . . . . . . . . . . . 56
6.2 Item content . . . . . . . . . . . . . . . . . . . . . . 57
6.3 Item format . . . . . . . . . . . . . . . . . . . . . . . 62
6.4 Information in item distractors . . . . . . . . . . . . . 63

7.0 RECOMMENDATIONS . . . . . . . . . . . . . . . . . . . . . . . 64

REFERENCES . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

APPENDIX A RESONING TEST . . . . . . . . . . . . . . . . . . . . 75

APPENDIX B PERCENTAGE OF TESTEES WHO SELECTED A PARTICULAR OPTION 104

# LIST OF TABLES

TABLE 4.1 CLASSIFICATION OF THE REASONING TEST ITEMS ACCORDING TO TYPE . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

TABLE 4.2 ABILITY GROUPS ACCORDING TO RANGE OF TOTAL SCORES ON THE SPM, AND NUMBER OF TESTEES IN EACH ABILITY GROUP . . . . 20

TABLE 5.1a SIGNIFICANCE OF THE DIFFERENCES (Z) BETWEEN THE PROPORTION CORRECT (P) OF THE DIFFERENT POPULATION GROUPS WITH REGARD TO CUES IN TEST ITEMS . . . . . . . . . . . . . . . . . . 23

TABLE 5.1b BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.1a . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

TABLE 5.2a ODDS RATIO, PROPORTION CORRECT(P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR IDENTICAL ITEMS AND THE SAME/DIFFERENT POSITIONS OF THE CORRECT ANSWERS . . . . . . . . . . . . . . 26

TABLE 5.2b BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.2a . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

TABLE 5.3 SIGNIFICANCE OF THE DIFFERENCES ($\chi^2$) BETWEEN POPULATION GROUPS IN THE CHOICE OF ITEM DISTRACTORS . . . . . . . . 28

TABLE 5.4a ODDS RATIO, PROPORTION CORRECT(P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$)WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR TRUE/FALSE ANSWER OPTIONS. . . . 30

TABLE 5.4b BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.4a . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

TABLE 5.5a ODDS RATIO, PROPORTION CORRECT(P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR POSITIVE/NEGATIVE TEST ITEMS . . 32

TABLE 5.5b BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.5a . . . . . . . . . . . . . . . . . . . . . . . . . . . 32

TABLE 5.6a ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR CONCRETE AND ABSTRACT SYLLOGISMS . . . . . . . . . . . . . . . . . . . . . . . 34

TABLE 5.6b BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.6a . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

TABLE 5.7a ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR LETTER AND NUMBER SERIES, VERBAL AND NON-VERBAL (FIGURE) ANALOGIES . . . . . . . . . . . . . . . . 38

TABLE 5.7b BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.7a . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

TABLE 5.8a ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR TIME CONCEPTS . . . . . . . . 42

TABLE 5.8b BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.8a . . . . . . . . . . . . . . . . . . . . . . . . . . . 43

TABLE 5.9a PROPORTION CORRECT (P), AND CONFIDENCE INTERVAL FOR P, WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR APPLIED AND THEORETICAL KNOWLEDGE OF COMPASS DIRECTIONS . . . . . . . 45

TABLE 5.9b BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.9a . . . . . . . . . . . . . . . . . . . . . . . . . . . 45

TABLE 5.10a ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR MATHEMATICAL "STORY" PROBLEMS AND CALCULATIONS . . . . . . . . . . . . . . . . . . . 47

TABLE 5.10b BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.10a . . . . . . . . . . . . . . . . . . . . . . . . . . 47

TABLE 5.11a PROPORTION CORRECT (P), AND CONFIDENCE INTERVAL FOR P, WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR STATEMENT-TYPE ITEMS AND "OTHER" VERBAL ITEMS . . . . . . . 49

TABLE 5.11b BIAS INDICES ($\chi_3^2$ AND ICC) FOR THE STATEMENT ITEMS IN TABLE 5.11a . . . . . . . . . . . . . . . . . . . . . . 49
TABLE 5.12a ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR ANTONYMS AND SYNONYMS . . . 51
TABLE 5.12b BIAS INDICES ($\chi_3^2$ AND ICC) FOR THE ITEMS IN TABLE 5.12a . . . . . . . . . . . . . . . . . . . . . . . . . . 51
TABLE 5.13a ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR FIGURE SERIES AND FIGURE ANALOGIES . . . . . . . . . . . . . . . . . . . . . . . . 53
TABLE 5.13b BIAS INDICES ($\chi_3^2$ AND ICC) FOR THE ITEMS IN TABLE 5.13a . . . . . . . . . . . . . . . . . . . . . . . . . . 53

## LIST OF FIGURES

FIGURE 5.1 GRAPHICAL REPRESENTATION OF THE MEAN ITEM DIFFICULTY VALUES FOR CONCRETE AND ABSTRACT SYLLOGISMS . . . . . . . . 36

## EKSERP

'n Belangrike vraag by itemsydigheid is nie alleen watter tipe items meer dikwels as ander tipes sydig is nie maar veral wat die oorsake van sodanige sydigheid is. In hierdie verslag word spesifiek aandag geskenk aan die rol wat iteminhoude en itemformate ten opsigte van itemsydigheid speel.

'n Redeneringstoets waarin die iteminhoude en -formate doelbewus gemanipuleer is ten einde inligting ten opsigte van hetsy sydigheid of differensiële groepprestasie te verskaf, is spesiaal vir hierdie ondersoek opgestel en op blanke, Indiër - en swart skoliere in st. 7 toegepas. Sydigheidsindekse is bereken met behulp van die Scheuneman-tegniek asook die drieparameter logistiese model (itemresponsteorie).

Op grond van die resultate van die ondersoek word 'n onderskeid gemaak tussen ware en skynitemsydigheid. Aspekte van ware itemsydigheid sluit onder andere in taalfaktore, té aanloklike itemafleiers en 'n gebrek aan kennis en begrippe wat hoofsaaklik op informele wyse verkry word. Skynitemsydigheid aan die anderkant spruit nie soseer uit die item as uit die toetsgedrag van die toetsling nie, byvoorbeeld 'n neiging om te klassifiseer op grond van assosiatiewe soortgelykheid (in plaas van konseptuele soortgelykheid), selektiewe aandagskenking en 'n gebrek aan 'n logiese ingesteldheid.

Wat iteminhoude en -formate betref, is bevind dat...

- items waarin 'n stelling gemaak word swakker funksioneer as ander tipes verbale items.
- die verhaaltipe inhoud tot besonder swak prestasie lei.
- gesegdes nie 'n geskikte itemtipe is nie.
- figure 'n beter iteminhoud as taal is.
- figuuranalogieë beter funksioneer as verbale analogieë.
- reekse 'n beter formaat as analogieë is.
- sillogismes (ongeag inhoud) nie 'n geskikte itemtipe is nie.

Weens die beperkte aantal items wat telkens ter sprake is, moet die bevindinge van die ondersoek as rigtingwysend en verkennend eerder as definitief beskou word.

# ABSTRACT

An important question with regard to item bias concerns not so much which questions show bias more frequently than others, but what the causes are of such bias. This report pays particular attention to the role played by item content and format in this regard.

This study involved the construction of a reasoning test in which the content and format of items were purposefully manipulated to obtain information regarding bias or, alternatively, differential group achievement. The test was administered to white, Indian and black pupils in standard 7. Bias indices were calculated by means of the Scheuneman technique and the three-parameter logistical model (item response theory).

The results of the study have led to a differentiation between true and apparent item bias. True item bias includes aspects such as language factors, overly attractive distractors, and a lack of knowledge and concepts that are usually informally acquired.

Apparent bias, on the other hand, is a feature of the testee's test behavior rather than of the item, for example, a tendency to use associative (rather than conceptual) similarity in classification, selective attention and an insufficiently logical approach.

As regards the content and format of items, it was found that ...

- items in which a statement is made function worse than other verbal item types.
- story-type items lead to particularly bad performance.
- sayings are not suitable item types.
- figures are better than language in terms of content.
- figural analogies function better than verbal analogies.
- series are better than analogies in terms of format, and
- syllogisms (regardless of content) are not suitable item types.

Since the number of each item type included in this study was limited, these results should not be seen as conclusive, but as exploratory in nature and suggestive of further directions of study.

## 1.0 INTRODUCTION

Research in South Africa regarding test bias is still in its infancy. There has been no particular need for such research since the tests compiled for the various ethnic groups to date have largely been separate. Changing political and industrial conditions in South Africa are now leading to situations that cannot be handled by separate tests, for example inter-group competition. However the use of joint or common tests poses a new problem, namely that of bias. The question can rightly be asked whether a group that was not included in the standardization process of a specific test will not be at a disadvantage when taking that test. If the intention is to introduce common tests for the various population groups, which comply with the demand for lack of bias (so that test scores reflect true and not artificial differences), attention will have to be given to the extent of bias in tests as well as to the nature, operation and causes of bias. Although it is important to know what types of items are often biased, it is even more important to determine why these items are biased. It is more important for the test compiler to promote insight and understanding than to identify and eliminate biased test items.

In previous reports (Owen 1986, 1989) attention was given to test bias (bias with respect to predictive and construct validity) as well as to the identification of biased items. In this report the focus shifts to potential sources or causes of item bias. Specific issues studied were the roles played by item content and item format in item bias.

## 2.0 AIM

The focus in this study was on those sources of bias located in the characteristics of the test item. Most investigations so far (excluding that by Scheuneman (1985)) were aimed at finding deviating items. Although these investigations produced valuable information regarding the "best" methods of identifying item bias, they left a gap with respect to systematic and objective knowledge about the *nature* of bias and the reasons *why* a specific item is biased. In the words of Ironson (1982:153):

> *We need to know more... about what makes an item biased.*

Despite the fact that Lord (1977) asked in the previous decade

> ... *why certain items are biased, why certain groups of people respond differently than others on certain items, and what can be done about this* (p. 29),

the problem remains largely unsolved. Rudner, Getson and Knight (1980:228) add the following:

> ... *future research should be directed toward the identification of content and characteristics which meaningfully contribute to bias in test items.*

Carlton and Marco (1982:304) put it even more strongly:

> *Identifying factors that explain why items are biased should be one major goal of test development research.*

The aim of the present investigation was to identify not only the characteristics and content that contribute meaningfully to bias in test items, but also - more importantly - the factors on the basis of which the reasons for bias can be determined. More specifically the intention was to vary the two essential characteristics of a test item, namely content and format, in a systematic manner in order to identify any differential effect that these two aspects may have on the performance of white and black testees. In this way it should be possible to explain bias on the basis of the manipulations carried out.

A shortcoming in most of the investigations into test bias has been the use of an *ex post facto* approach. In other words, after an item had been administered and bias had been identified, an attempt was made to conclude from a study of the item why the item was biased. According to Schmeiser (1982) an experimental design is required in which the item format (independent variable) is manipulated. This concept was put into practice in this investigation.

## 3.0 ITEM BIAS

To understand why various groups - that differ considerably with regard to culture and other matters - perform differently with respect to a specific item it is necessary to keep certain background factors in mind. Although a large number of such factors can be distinguished, for instance the quality of school education, motivation, test anxiety and test sophistication, only four of the most important ones will be briefly discussed here, namely culture, socio-economic status, language and cognitive style.

## 3.1 SOME BACKGROUND FACTORS WITH REGARD TO ITEM BIAS

### 3.1.1 Culture

The reference to *cultural differences* is probably the most common criticism levelled at standardized tests. For example, the Council of the Society for the Psychological Study of Social Issues says the following:

> *We must also recognize the limitations of present day intelligence tests. Largely developed and standardized on white, middle-class children, these tests tend to be biased against black children to an unknown degree* (quoted by Jensen 1974:190).

A specific culture stimulates a specific form of cognitive development, in other words intellectual ability is *linked to culture* (Lesser, Fifer & Clark 1965; Scarr 1981; Anastasi 1970; Sundberg & Gonzales 1981). According to Lesser *et al.* (1965:3) the study of *cultural differences* with regard to intellectual ability is becoming a fundamental issue in education and psychology:

> *How do we provide valid psychological evaluation of children from widely dissimilar groups?*

One reason why no definite answers have yet been found to this question is, according to Anastasi (in Lesser *et al.* 1965:11), the fact that a concept such as intelligence is based on a Western technological culture:

> *It is not so much that tests are unfair to lower-status groups, as that lower-class environment is not conducive to the effective development of 'intelligence' as defined in our culture.*

Various researchers agree that every culture encourages and promotes the development of certain abilities or types of behaviour while discouraging and suppressing others (Anastasi 1976). The possible relation between culture and cognitive development becomes clear in Ferguson's (1954) prerequisite for the development of an ability: The culture must provide opportunities for "overlearning" an activity. It follows that the absence of a certain activity in a culture is accompanied by a related absence in ability. Irvine(1970) points out that the traditionally oral teaching of skills in certain non-Western communities gives rise to rules for intelligent behaviour that are unique to those communities. Their behavioural applications of knowledge may also be completely different from those of the individualized, competitive and industrialized West. Furthermore, rural people with little schooling seem unable to accept and draw conclusions from the logical assumptions of syllogisms (Cole & Scribner 1974). However it is clear from an experiment described by Brislin (1983) that black pupils are able to work out syllogisms that are formulated in their mother tongue.

According to Triandis and Brislin (1984) the relation between culture and cognitive development can be viewed from three perspectives, namely the *universal* (all cultures show certain similarities with respect to cognition), the *evolutionary* (differences in cognitive functioning are sought in the activities of the group concerned) and the *relativistic* (the focus falls on differences between cultural groups). The above writers maintain that all three approaches are valid to a certain extent. For example, the existence of universal characteristics cannot be denied: all people categorize, use opposites and classify. However several writers point out that, although the components of the cognitive system, in other words memory, categories, associations, syllogisms, coding and decoding, semantic integration and verbal explanation, are found in most cultures, they are related in an extremely complex way and that variations occur as a result of the specific characteristics of the culture concerned (for instance literacy) (Triandis & Brislin 1984; Bond 1981; Irvine 1970). In this regard Anastasi (1985:122) states the following:

> *These differences are reflected in the development of abstract thinking and in the nature and breadth of the concepts that are formed.*

Ebel's (1979) statement that what a person learns depends on the culture in which he finds himself is illustrated very well by Medley and Quirk (1972, 1974). These writers conclude that if the 65 items comprising the Common Examinations of the National Teacher Examination were to be replaced by "black" items (i.e. items that deal with the activities of

American Negroes in social life, literature and the arts, but that are not so esoteric that an informed person from another population group cannot answer them too), blacks would improve their test performance by 27 % in relation to that of the whites. These findings correspond with those of Nancy Abrams (1979) who found that blacks performed about one and a half standard deviations better than whites in a "black" test.

The effect of culture on intellectual ability can take many forms. In addition to direct learning situations in and out of school the typical behaviour code of the community also has a subtle influence. Guthrie (1963) posed the question whether members of an authoritarian and traditional community develop a "differently" organized reasoning ability. Almost twenty years later Ghuman (1980) found indications that this could be the case. According to him Punjabi youngsters displayed a convergent style of thought that could be ascribed to the rigid form of authority maintained by the teachers in their schools and encouraged by the community as a whole.

To summarize it can be said that cognitive development and the acquisition of certain abilities are to a great extent culture-related. On the other hand the statement by Buss (1977) may also be true, namely that cultural differences are not a matter of differences in skills per se but in the situations in which skills are applied. This view corresponds with that of certain modern anthropologists:

> *...the belief systems and cultural premises of traditional people may differ from those in industrialized societies, but they embody the same logical processes and concern with relation of cause and effect* (Cole & Scribner 1974:25).

It is clear from the above that although the influence of culture is extremely subtle in many ways and difficult to observe directly, it may still be an important source of item bias, particularly in plural communities with divergent cultural backgrounds.

### 3.1.2 Socio-economic status (SES)

There is apparently little doubt that socio-economic circumstances affect test performance in one way or another, that better circumstances are to the advantage of certain individuals and groups when taking an intelligence test, that children of a higher social status generally perform better than children of a lower social status (Havighurst & Breese 1947; Lorge 1952-53). In recent years the question of the role of socio-economic factors in test bias has become very important to psychologists (Oakland & Feigenbaum 1979).

What are the characteristics of an underprivileged person? According to Novick (1980) our knowledge of what constitutes a disadvantage or handicap is still incomplete. However certain factors related to educational disadvantage can be identified, for example poverty, health, malnutrition, inadequate school facilities, maladjustment, absence of cultural stimulation, illiterate and superstitious parents, lack of books, pictures, furniture, toys, television, and all the other equipment of modern society that contribute to shaping the thoughts of the

Western child, lack of discipline and particularly lack of a stable relationship with a supporting adult, a feeling of hostility and helplessness with respect to the social and educational establishment (Vernon 1968; Scarr 1981; Thorndike & Hagen 1969). The cultural level at home, the parents' level of education and the extent to which they encourage their children are some of the most important factors that influence the performance of children in intelligence and verbal tests (Radford & Burton 1972; Vernon 1968; Eells *et al*. 1951).

Children from underprivileged areas often perform poorly in psychological tests because these tasks are strange to them and are regarded as unimportant (Lesser, Fifer & Clark 1965). Improvements in a child's environment may lead to improved test performance (Walberg & Haertel 1984). There is little doubt that the IQs of schoolchildren correlate with the socio-economic status of their parents (Jensen 1969; Humphreys & Taber 1973). Wigdor and Garner (1982) estimate the correlation between SES and ability test scores to be approximately 0,30. The magnitude of the differences in performance between low and high SES groups is clear from the fact that the average test scores of pupils from families in the top 20 % of the socio-economic structure lie on approximately the 65th percentile relative to the general population, while the average scores of children whose families fall into the bottom 20 % lie approximately on the 35th percentile (Wigdor & Garner 1982).

According to Jensen (1969) children with a low SES perform equally well in Jensen's Level I abilities (in other words with respect to associative processes such as memory) regardless of their ethnicity. Further to this Horn (1976) found that low and high SES groups performed equally well in Level I abilities. MacKenzie (1981) has however found a stronger relationship between SES and Level abilities (in other words abilities involving conceptual and reasoning processes).

The way in which the different SES groups approach and solve problems in test items is instructive, not only with regard to certain differences between the groups, but also with respect to a better understanding of the possible causes of item bias. The following are some of the most important findings with regard to the relation between SES and performance in test items:

- Children of a low SES complete the test more quickly and often select a distractor at random if they do not understand the principle concerned, for instance series completion (Wolf 1978; Eells *et al*. 1951). Similar observations were made among Puerto Rican schoolchildren in New York and children in Hawaii (Marmorale & Brown 1979).

- If the structure of the item is complicated, for example in numerical problems, children of a low SES find it more difficult (Wolf 1978).

- The incorrect answer that corresponds most closely to the correct one is almost equally popular among low and high SES groups. However the low SES group more often selects less probable answers (Wolf 1978) and their answers are also distributed more uniformly across all the possible answers or distractors (Eells *et al*. 1951).

- Certain item types such as analogies, opposites and classification can be revised or rewritten more easily than others, such as syllogisms, to eliminate the effect of middle-class social status on performance (Haggard 1954).

- The mean difference in status is largest for verbal items and smallest for items comprising pictures and geometric figures (Eells *et al.* 1951).

- Almost all items associated with unusually large status differences consist of verbal symbolism; a large number of these items involve academic or bookish vocabulary (Eells *et al.* 1951).

- Items associated with small status differences almost without exception consist of either non-verbal symbolism or simple, common words (Eells *et al.* 1951).

- Motivated children of a low SES perform significantly better than non-motivated children of the same status (Haggard 1954).

- If the aim is to narrow the considerable gap in performance between low and high SES groups, it is necessary in the first place not only to increase the lower status group's familiarity with the type of material concerned, but also to increase their motivation to perform well (Haggard 1954).

In summary it can be said that socio-economic status is related to the type of item, the degree of difficulty of the item, the typical response pattern to the item, familiarity with the type of material comprising the item and the motivation to perform well. All these points must be remembered when compiling tests for use with different SES groups.

### 3.1.3 Language

As concerns the language used in a test Oakland (1977) points out that it should be of such a nature that every testee has the best opportunity to understand what is expected of him and to react freely and with ease. If this is not the case the language in the test can contribute to cultural bias (Schultz & Fortune 1981:119) Although language is extremely important - in many cases it is responsible for the disparity between verbal and non-verbal scores - it becomes less significant as the time spent on testee instruction, using the test medium, increases (Jensen 1980).

The translation of a test gives rise to many problems with respect to concepts, cultural interpretation and connotations (Merryfield 1985; Olmedo 1981). Translation into non-standard dialects is no solution and leads only to minor improvements in the performance of black children (Oakland 1977). Elements in the test material such as a word or a sentence, a diagram or a figure, can stimulate different associations in different cultures (Vernon 1969).

When a bilingual testee takes a test that is not presented in his dominant language - which is often the case in Africa - a test that is supposed to measure intelligence, for example, could also measure aspects it is not supposed to measure such as language proficiency (Guthrie 1963; Hale 1982).

Black Africans have a unique style of language, for instance they will give an elaborate description rather than an exact definition of a certain matter. They regard a direct statement or approach as rough and unimaginative, while camouflage by constantly changing definitions is regarded as intelligent behaviour (Hale 1982). It is not yet clear at this stage to what extent this traditional language usage is found among black primary and high school pupils in South Africa too - owing to the influence of Western-type school instruction it is suspected that it does not play a significant role in test scores.

Black pupils in South Africa receive instruction in English from Standard 3. They can therefore be expected to have a reasonable knowledge of the language by the time they reach high school. If it is also taken into account that test compilers try to limit the role of specific language skills to a minimum in tests that do not specifically measure such skills (e.g. verbal reasoning tests), it can be assumed that in this investigation language does not have a fundamental effect on the test scores on non-language tests. However if the language skills of testees are poorer than expected, these skills can play a role in verbal reasoning as well.

### 3.1.4 Cognitive style

Research findings reveal that the typical cognitive style of individuals could contribute to group differences in test performance and therefore to test and item bias as well. Of all the identified types of cognitive style (see for example Coop & Sigel 1971; Guilford 1980; Federico 1983), the bipolar field-dependent (FD)/field-independent (FI) style is not only the most well researched, but also has the widest application in psychology and education (Templer 1972; Witkin *et al*. 1967, 1977).

Field-dependence and field-independence (FD and FI) refer to an individual's tendency to organize his experiences in an analytical or a global way. An FI person is able to focus on a specific stimulus regardless of the presence of a number of attractive yet irrelevant stimuli. This tendency is more than simply an approach, it represents a general analytical orientation. An FI is able to overcome or restructure the organization or structure of a given field. An FD accepts the field as it is and finds it difficult to distinguish the parts of the field from the overall context. This global tendency indicates a limitation with regard to differentiation.

As far as the development of this type of cognitive style is concerned it seems that individuals who grew up in a climate emphasizing autonomy and performance tend to be field-independent, while those who grew up in a protective and conforming environment tend to be more field-dependent. On the basis of this assumption whites can therefore be expected to be more field-independent and blacks more field-dependent.

In general FD testees have been found to perform worse than FI testees (Globerson *et al.* 1985), when misleading cues to the solution are given. However in the absence of such misleading information, FD testees have performed just as well and even better than FI testees (Witkin *et al.* 1977). According to the latter writers FI testees performed better than FD testees in mathematics, science, engineering and architecture. Kaufman (1979) points out that FI testees performed considerably better than FD testees in laboratory tests for spatial orientation and in psychometric tests such as hidden figures, block design, etc.

The incidence of a certain type of cognitive style is not limited to a specific ethnic or socio-economic group. In the USA it was found that blacks and other low-income groups tended to use the relation style, but so did higher-income groups (Hale 1982). Although no group used a certain cognitive style exclusively, a definite relationship was found between style and cultural or ethnic group, particularly when the group represented a specific point on the socio-economic scale. Hilliard (quoted by Hale 1982:42) makes the following claims in this regard:

- Afro-Americans tend to react to objects in terms of the total image and not the various parts, and Euro-Americans tend to divide everything into smaller parts which together form a whole.
- Afro-Americans tend to prefer inference reasoning to deductive and inductive reasoning.
- Afro-Americans tend to use estimates rather than actual values when working with space, numbers and time.
- Afro-Americans tend to focus on people and their activities rather than on objects.

Several researchers (for example Kaufman 1979; Federico 1980; Hunt 1980; Rohrl 1979; McGee 1979) point to a relationship between cognitive style and the functioning or activities of the left and right hemispheres of the brain. Among normal right-handed people the functioning of the left hemisphere is associated with a verbal, analytical, sequential syllogistic and objective way of processing information, while that of the right hemisphere is associated with a spatial, synthetic, nonverbal, creative, holistic and intuitive method of processing information.

According to Kaufman (1979:97) the basic difference between the left and the right hemispheres lies not so much in the stimuli that are processed, but rather in the way in which they are processed:

> *The left hemisphere specializes in sequential* (his underscoring) *processing, analytic, linear, and successive in nature; the right hemisphere is lateralized for simultaneous* (his underscoring) *or multiple processing, holistic in nature.*

This means that each hemisphere of the brain has its own rules for processing information and that the brain therefore has two "intelligences" that function independently (in view of this Witelson (1977) formulated the hypothesis that persons suffering from dyslexia have two right hemispheres). The left hemisphere, which represents the scientific method, organizes data according to the principle of conceptual similarity, while the right hemisphere functions according to the principle of structural similarity. Kaufman (1979) illustrates the above with the aid of an example taken from the Stanford-Binet Intelligence Test: The left hemisphere regards an apple and a pear as similar, because both are types of fruit, but to the right hemisphere they are similar in being round.

During the past decade psychological theories regarding the functions of the left and the right hemispheres of the brain came to the attention of anthropologists, who made certain interesting observations. One of these was that left-hemispherical functions increase in importance as communities became dependent on artifactual recording devices such as measuring sticks and scales and calibrated monitoring equipment such as watches and calendars (Rohrl 1979:356).

Ethnic differences in test performances are apparently related to the functions of the left and the right hemispheres of the brain. Although individuals use both hemispheres, it appears that cultural and other factors may promote the development of a certain cognitive or thinking style. In an attemp to describe the connection between cultural factors and cognitive style, Rohrl (1979) related certain cultural factors to brain functions.

The cultural correlates of the functions of the left hemisphere include:
- Specialized social organization
- Extensive dependence on verbal communication
- Large and impersonal community
- Individualism
- Day-to-day pattern of existence not directly dependent upon nature
- Much verbal and formal instruction
- Dependence on watches, calendars, etc.

Cultural correlates of functions of the right hemisphere include:
- Shared-function social organization
- Relatively little need for verbal communication
- Small, open (face-to-face) community
- Tradition and respect for elders and their advice
- Closeness to nature
- Learning through example rather than verbal instruction
- Relatively little need for watches, calendars, etc.

According to Rohrl (1979) the presence of a number of variables from one of these two cultural sets indicates a preference for the corresponding brain function. Although the above cultural correlates have not yet been confirmed, Kaufman (1979:103) maintains that ...*research suggests a possible right hemisphere leaning among blacks*. The latter writer's assumption is supported by the fact that blacks excel in typical right hemisphere functions as measured in performance in figure completion tests and visual spatial tasks such as pattern recognition. Blacks also have considerable talent associated with functions of the right hemisphere, for example music, creative dancing and visual art.

Some of the cultural correlates identified by Rohrl (1979) with respect to right hemisphere functions certainly apply to a large number of blacks in South Africa (see for example Van der Vliet 1974; Dubb 1974). However this does not necessarily mean that South African blacks have a greater tendency to right hemisphere functions than whites. If this is the case, it may contribute to group differences in test performance and be a cause of bias.

It is clear from the above discussion that cognitive style is an additional viewpoint from which group differences in test performance can be studied. It has been indicated that cognitive style can affect the test performance of individuals and groups in a variety of ways, depending on the type of test being used, the intellectual tasks required and the specific hemispherical brain function involved. The above also has certain implications in terms of the causes of item bias. To mention only a few: Black pupils can be expected to perform better in concrete than in abstract test items, or they can be expected to experience relatively more difficulty with test items that contain time concepts.

## 3.2 DEFINITIONS OF ITEM BIAS

Investigations into item bias are aimed at determining whether different ethnic, racial or cultural groups display different behaviour patterns with respect to test items (Mellenbergh 1983). A typical statistical indication that a test item may not be appropriate or suitable for a specific cultural group is that the item is obviously too difficult (or too easy) for the group concerned (Rudner 1978; Thorndike 1982). According to one definition an item is regarded as biased if it favours one group unfairly over another group (Wilcox 1984/85), or if it systematically functions differently for different groups (Plake & Hoover 1979/80; Burrill & Wilson 1980).

According to Cleary and Hilton (1968:61) an item is biased for members of a group

> *...if, on that item, the members of the group obtained an average score which differs from the average score of other groups by more or less than expected from performance on other items of the same test.*

According to Cleary and Hilton's definition item bias can be identified only on the basis of the relation between the item concerned and other items in the test. Further to this Shepard, Camilli and Averill (1981:317) state that methods for disclosing item bias

> *... identify deviant or anomalous items in the context of other items.*

Cleary and Hilton's definition is based on features of classical test theory and is unconditional with respect to assumptions about the ability level of the testees. However more recent definitions of item bias are increasingly based on item response theory (latent trait theory) (Humphreys 1986). An important feature of these definitions is that a precondition is set with regard to the testees' ability level: only testees with the same level are compared with one another (Mellenbergh

1983). The following is a typical definition of item bias based on the item response theory:

> ... *an item is generally considered biased if equally able members of different groups have unequal chances of success on the item* (Subkoviak *et al.* 1984:49).

This view of item bias is presently shared by most researchers, for example Petersen (1980), Scheuneman (1980, 1981), Marascuilo and Slaughter (1981) and Kulick and Dorans (1984).

## 3.3 BIASED ITEM TYPES

The aim of research into bias is not only to provide guidelines and procedures for test compilers to follow in identifying and eliminating apparently biased test items, but particularly to identify variables or factors that may be responsible for bias with respect to specific groups (Schmeiser 1982). So far two main approaches have been followed in studying the problem of item bias, namely the judgmental approach (Tittle 1982) and the statistical approach (Angoff 1982; Ironson 1982, and others). Many test manufacturers in the USA, such as the Educational Testing Service and the Psychological Corporation, use both methods to identify biased items (Carlton & Marco 1982; Coffman 1982; Green 1982; Lenke 1982). A general lack of agreement between the statistical and the subjective evaluation of item bias (Plake 1980) creates a particular problem, and experts are often unable to explain what causes bias (Shepard 1981; Shepard, Camilli & Williams 1983; Reschly 1984). The assumption made by Burrill (1981:143) in this regard represents a typical finding: *The item data can show how bias occurs but not why*. The cardinal question is therefore: What causes an item to be biased for a specific group?

In order to identify the variables that are potentially responsible for item bias it is first necessary to determine whether bias can be associated with a certain type of item - with regard to content as well as format (Linn & Harnisch 1981). Although little has been said on this matter (Stricker 1982), certain trends have been identified that can serve as basis for generating hypotheses regarding item bias.

### 3.3.1 Item content

- According to Scheuneman (1978:6) the main source of item bias is content ... *which presupposes experience or knowledge not equally available to members of different groups for reasons of cultural background or economic disadvantage*. However Jensen (1980) maintains that it plays a relatively minor role. The removal of such items is also not always advisable because it could affect the predictive validity of the test, for instance in the case of a language test for Puerto Ricans in the USA. A reasonably consistent finding is that blacks find items containing material relevant to their culture, easier than others (Breland *et al.* 1974; Medley & Quirk 1974; Abrams 1979).

- Information in the item. On the basis of the knowledge that blacks generally have a less well-developed vocabulary than whites, Scheuneman (1983) speculates that white testees tend more than black testees to use the information contained in an item to select a correct response, whether through elimination of some of the distractors or by means of another strategy. Handrick and Loyd (1982) as well as Plake and Huntley (1984) found that some pupils react to subtle grammatical cues in the item stem and distractors when answering an item.

- Negatively phrased items. Scheuneman (1980) analysed more than 2 000 items of the Metropolitan Readiness Tests and the Otis-Lennon School Ability Test and found that blacks apparently performed less well when the stem of the item was phrased negatively or when one of the possibilities was FALSE or something similar. However Burrill (1981) and Shepard, Camilli and Williams (1983) maintain that Scheuneman's finding is the exception rather than the rule. That there is more to Scheuneman's findings than some researchers suspect is clear from the work by Dudycha and Carpenter (1973:120). They found that negatively phrased items were more difficult than positively phrased items, perhaps because ...*the negative stem is a departure from expectation and requires a shift of mental set, which test takers fail to do*. They also found that items where one of the distractors is comprehensive (for example "none of these") are more difficult than items in which the distractors are specific - perhaps because an item with a comprehensive distractor demands more knowledge from the testee: not only does he have to know what the correct answer is, but also which answers are wrong. The latter finding is supported by Hughes and Trimble (1965). However the inclusion or omission of the category "none of these" does not affect the reliability or validity of the test (Williamson & Hopkins 1967).

- Mathematical problems, particularly verbally formulated mathematical problems, and the application of mathematical concepts often appear to be biased with respect to blacks (Fishbein 1975; Forsyth & Spratt 1980; Linn & Harnisch 1981; Green 1982; Shepard, Camilli & Williams 1983), perhaps because this type of item requires more than one ability to solve it correctly (Reckase 1985). Simple mathematical problems such as determining the square root of an integer are extremely difficult for minority groups (this type of knowledge is obtained mainly at school). On the other hand, items requiring everyday mathematical knowlegde such as counting money, are experienced as relatively easy (Breland *et al.* 1974; Ironson & Subkoviak 1979). In general it appears that blacks perform less well than whites with respect to quantitative material (Scheuneman 1979). Mistakes that are made (by all groups) are often systematic, for instance the consistent use of a multiplication sign as if it denotes summation (Tatsuoka & Linn 1983).

In light of the above findings (with the exception of certain tendencies), one can agree with the statement by Linn *et al.* (1981:159) that a comparison between a biased and an unbiased item with respect to content ... *did not lead to the identification of any systematic content differences*. For this reason Rudner (1978) prefers to classify the reasons for item bias on the basis of the item construction rather than content.

### 3.3.2 Item format

- The format in which an item is presented, i.e. relatively independent of content and context, is another possible cause of bias. However such a source of bias cannot be traced unless more than one type of item format is used to measure the same underlying ability. According to Frederiksen (1984) modifying the format of an item, for example from a multiple-choice question to an open-ended type, has only a minor effect on the construct being measured. However such changes can make an item more difficult. In this regard Dudycha and Carpenter (1973) found that items with an "open" stem (for example "steam is a form of...") are more difficult than those with a closed stem. No explanation could be found for this phenomenon.

- Syllogisms. According to Hunt (1982) reasoning/thinking can be described as a set of intellectual skills linked to specific content areas. For example persons who are able to reason on the basis of syllogisms with familiar content are often unable to answer a similar syllogism phrased in an abstract manner. This observation was confirmed by cross-cultural research. Haggard (1954) also points out that of the majority of item types syllogisms are the most difficult to rephrase in order to decrease "middle-class bias", as this type of problem is to a great extent academic, artificial and completely alien to the practical world of the lower-class child.

- Verbal items: synonyms, antonyms and analogies. Of the three commonly used verbal item formats, antonyms and analogies have the higher incidence of bias with respect to blacks, (Scheuneman 1980, 1981; Kulick & Dorans 1984), and also often lead to language problems and differences (Valencia & Rankin 1985). According to Scheuneman (1980:147) blacks appear to have difficulty with comparative terms such as "fewer", "closer" and "larger". This probably indicates certain language deficiencies on their part. In this regard Bond (1981) points out that it is impossible to separate vocabulary from verbal-analogical reasoning although the two differ conceptually. As far as sexual differences are concerned it was found that antonyms are more often biased with respect to girls (Donlon, Hicks & Wallmark 1980; Scheuneman 1981); antonyms and analogies with practical or scientific content favour men, while women are favoured if these items concern human relations (Stricker 1982).

- Non-verbal. Non-verbal reasoning items (excluding quantitative item types) usually consist of figures and other geometric designs that are presented in the form of series, analogies ("-is to-"), matrices, etc. If it is difficult to determine the cause of bias in verbal items (see for example Burrill 1981; Raju 1982), it is even more difficult in the case of non-verbal items where causes are even less "visible". In addition to this problem there is a lack of empirical findings with regard to specific biased item types, and the findings by Scheuneman (1981) are almost the only ones that have been reported. According to her, figure series are often more biased than figures presented in a different format. Bias is also more often found in letter series than in other types of series.

To summarize, ... *we still have an imperfect idea of the nature of item bias* (Metropolitan Achievement Tests 1978:8). More knowledge is needed not only with respect to the causes of item bias, why one group performs very well and another very poorly in an item, but also (as recommended by Nagy and Griffiths (1982) and Haney (1984)) with respect to how the individual actually experiences the task required of him by an item. Cognitive psychologists such as Pellegrino and Glaser (1980), Snow (1980) and others who study the processes underlying responses to items could perhaps provide more clarity in the future. In any event Scheuneman (1982) is confident that research with regard to item bias will eventually lead to a better understanding of bias and how it functions. However she also believes - and this is extremely important - that no-one should expect to find a single cause of bias on the basis of which a great deal of the average test differences between various racial groups can be explained (Scheuneman 1981).

## 3.4 HYPOTHESIS WITH REGARD TO ITEM BIAS

On the basis of the findings, cues and suggestions discussed in the preceding sections and the variety of item contents and formats that are available in the case of cognitive tests, the following hypotheses are formulated:

### Hypothesis 1

Different population groups do not derive equal benefit from the unintentional cues that items sometimes provide.

### Hypothesis 2

The position of the correct answer among the distractors does not affect the performance of the different population groups in the same way.

### Hypothesis 3

The different population groups do not react in the same way to the attractiveness of a specific distractor.

### Hypothesis 4

Population groups differ in their reactions to specific answer types - one false (untrue) answer, or one true answer.

### Hypothesis 5

The various population groups differ from one another in their reactions to positively and negatively phrased items.

Hypothesis 6

(a) Concrete and abstract items do not have the same effect on test performance, and
(b) the various population groups differ in their reaction to concrete and abstract items.

Hypothesis 7

Black pupils perform better in concrete than in abstract items.

Hypothesis 8

The various population groups differ from one another with regard to the effect of verbal and non-verbal (numbers and figures) items on their performance.

Hypothesis 9

Black pupils perform worse than white pupils in test items involving time concepts.

Hypothesis 10

Black pupils perform worse than white ones in items that involve direction and orientation with regard to direction.

Hypothesis 11

Black pupils find mathematical "story" problems more difficult to deal with than performing the actual calculations concerned.

Hypothesis 12

Black pupils perform worse in test items in which statements are made than in other types of verbal items.

Hypothesis 13

Black pupils find synonyms easier than antonyms.

Hypothesis 14

The figure series and figure analogy item formats
(a) affect test performance in different ways, and
(b) have different effects on the performance of the various population groups (the groups differ in their reaction to each member of an item pair).

## 4.0 RESEARCH DESIGN

### 4.1 SAMPLES

The testees comprised 1 056 white, 1 063 Indian and 1 093 black pupils in Standard 7. A description of the three samples appears in a previous report (Owen 1989). In the case of the white testees 20 schools in the Pretoria-Witwatersrand-Vereeniging (PWV) area were drawn at random in proportion to certain strata (Afrikaans-medium, English-medium and Arikaans/English-medium schools). In the Cape Peninsula 10 schools were drawn in a similar manner. In the case of the Indian pupils 30 high schools were selected at random from the list of high schools in and around Durban. In respect of the black pupils 30 schools were drawn at random from the list of high schools in the PWV area. Owing to the unrest situation the test programme could be carried out at only three of these schools and was consequently moved to Natal. The number of schools from the various areas in KwaZulu were the following: eight schools from the Pietermaritzburg area, eight from Empangeni and nine from Newcastle. Testing in the white and Indian schools took place during August/September 1985 and in the black schools in KwaZulu in June 1986. The average age of the white and the Indian pupils was 15,0 years and that of the black pupils 16,5 years.

### 4.2 MEASURING INSTRUMENTS

Four measuring instruments were used in the investigation, namely the Junior Aptitude Tests (JAT), the Raven Standard Progressive Matrices, the Reasoning Test and a biographical questionnaire. The JAT and the Raven have previously been discussed elsewhere (Owen 1989). The present report concerns the results obtained with the aid of the Reasoning Test.

#### Reasoning Test

A Reasoning Test comprising 85 items was compiled to test the hypotheses in Paragraph 3.4. This test was specifically concerned with item content and item format, in other words the essential characteristics of an item that could lead to bias.

It is obviously not possible to cover all item contents and formats within the scope of one investigation. Certain choices had to be made, for instance to include in the Reasoning Test only those contents and formats that appear in typical aptitude and intelligence tests. Even this restriction leaves room for a wide variety of possibilities, so further choices had to be made. If fewer item types are evaluated, more items can be created for each type and more valid conclusions can be drawn. On the other hand, the larger the number of item types that are evaluated, the fewer the items available for evaluation and the more uncertain the conclusions. Because this investigation was mainly of an exploratory nature, the latter method was selected.

The items in the Reasoning Test were constructed following a study of a large variety of item formats and contents used in modern aptitude, intelligence and other tests for measuring verbal and non-verbal reasoning ability. The following item formats appear to be commonly used: analogies, inferences, classifications, numerical calculations, matrices, series, syllogisms and linear syllogisms (for example "John is taller than Peter..."), statements and mathematical "story" problems. As regards item content, three basic elements, namely language, figures and numbers are used to create concrete and abstract, and meaningful and meaningless items. Items can be phrased positively or negatively, and answer types are numerous (such as open/closed, true/false, "all"/"none of these").

Reasoning Test items were constructed on the basis of various content-format combinations (as discussed above) and on the basis of item types known to be problematic for various population groups (cf.Par. 3.3). The items were constructed in pairs, with for example one pair member being a verbal analogy and the other a figure analogy, but with both requiring the same strategy for their solution. In this example item format (analogy) remained constant and content (language as opposed to figures) was manipulated. Before the test booklet was compiled the items were shuffled so that one version of the question was not answered directly after the other version of the same question. The Reasoning Test appears in Appendix A. The types of items comprising the test appear in Table 4.1. Four items from the JAT 2 (Numbers 15, 17, 18 and 21) were also included in the test (as Numbers 23, 33, 43 and 52) in order to study the effect of manipulating the position of the correct answer on the difficulty value of the item. In addition Items 1, 9 and 18 of the Reasoning Test were antonyms of certain items (Numbers 14, 20 and 22 respectively) in the JAT 4 (Synonyms).

## 4.3 PROCEDURES, METHODS AND TECHNIQUES

The test procedures were discussed in a previous report (Owen 1989). The test programme was conducted in 1985 at white and Indian schools and in 1986 in most of the black schools, and took two days per school to administer. On the first day the JAT and the Biographical Questionnaire were administered and on the second day the Reasoning Test and the Raven Standard Progressive Matrices. As mentioned before, the present report mainly concerns the items in the Reasoning Test.

The following methods and techniques were used to analyse the data (responses to the items in the Reasoning Test):

- Item analysis. The item analysis programme concerned provides information on, among other things, the difficulty value (p value) of the items (in other words the ratio of the number of pupils who answered the item correctly to the number of pupils who attempted the item), as well as the number and percentage of responses to the item distractors (the number of testees who selected a specific distractor - these results appear in Appendix B).

- Odds ratio. With the aid of this ratio (cf. 5.1.2) one can determine whether the manipulation of an item pair had any effect on the p value of the items concerned.

**TABLE 4.1**

**CLASSIFICATION OF THE REASONING TEST ITEMS ACCORDING TO TYPE**

| Item types | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Vocabulary (antonyms/ verbal comprehension) | | General reasoning (inferences/ linear syllogisms) | Statements | Syllogisms | Series (number/ letter/ figure) | Verbal analogies | Figure analogies | Computations | Direction-orientation |
| Item numbers | 1# | 3 | 4 | 5 | 2 | 6 | 15 | 7 | 47 |
| | 8 | 11 | 21 | 14 | 10 | 25 | 29 | 16 | 56 |
| | 9# | 12 | 41 | 24 | 19 | 27 | 35 | 46 | 71 |
| | 13 | 20 | 60 | 34 | 26 | 43* | 37 | 55 | 84 |
| | 17 | 30 | 78 | 44 | 28 | 45 | 49 | 63 | |
| | 18# | 40 | | 53 | 36 | 52* | 54 | 64 | |
| | 22 | 50 | | 61 | 38 | 62 | 66 | 70 | |
| | 23* | 59 | | 68 | 39 | 75 | 69 | 85 | |
| | 31 | 67 | | 74 | 48 | 76 | 82 | | |
| | 32 | 73 | | 79 | 57 | 80 | | | |
| | 33* | | | 81 | 58 | | | | |
| | 42 | | | 83 | 65 | | | | |
| | 51 | | | | 72 | | | | |
| | | | | | 77 | | | | |
| Total | 13 | 10 | 5 | 12 | 14 | 10 | 9 | 8 | 4 |

Note: * These items were taken from the JAT 2 (Reasoning).
# The antonyms of items in the JAT 4 (Synonyms).

- Confidence intervals and $\chi^2$ tests with respect to p values. On the basis of these values it can be determined whether the manipulation of the item pair had a statistically significant effect on the p values of the items (cf. 5.1.2).

Two methods were used to calculate the bias indices, namely Scheuneman's chi-square technique and the three-parameter logistical model (item response theory).

### The Scheuneman chi-square method

This method is explained in detail in Scheuneman (1980) and in Owen (1986, 1989). Briefly it amounts to dividing the testees of two or more groups into a number of ability categories (three, four or five) on the basis of their total test scores. Within each of these categories, a distribution of the correct responses for every item is obtained for each population group. The number of correct responses within each ability category represents the observed frequency in the chi-square test. The expected frequency is determined with the aid of the proportion of correct responses of all the testees, irrespective of group membership, who obtained scores within a specific ability category. In other words the expected value for a single cell ($E_{ij}$) is obtained by multiplying the proportion of testees whose total score fell within interval j and who answered the item correctly by the number of testees in the cell. This is

$$E_{ij} = \frac{O._j}{N._j} N_{ij}$$

where $O._j$ = the number of testees in the total score interval (ability category) j who answered the item correctly;
$N._j$ = the total number of testees in the score interval j, and
$N_{ij}$ = the total number of testees in group i and score interval j.

The test statistic ( $\chi^2$) is

$$\chi^2 = \Sigma \frac{(B_e - B_o)^2}{B_e} + \Sigma \frac{(W_e - W_o)^2}{W_e}$$

where $B_e$ = the expected frequencies of Group B;
$B_o$ = the observed frequencies of Group B;
$W_e$ = the expected frequencies of Group W; and
$W_o$ = the observed frequencies of Group W.

In this investigation of item bias in the Reasoning Test the ability categories were based on the testees' performance in the Raven SPM. The reason for using SPM rather than Reasoning Test scores was that the former were assumed to provide a more accurate indication of reasoning ability than the latter which is naturally made up of experimental items.

Based on the score distributions of the three population groups in the SPM four ability groups were distinguished. The distribution of SPM scores with respect to the ability groups, and the number of testees in each group are given in Table 4.2. This distribution of ability groups applies to all the items in the Reasoning Test.

TABLE 4.2

ABILITY GROUPS ACCORDING TO RANGE OF TOTAL SCORES ON THE SPM, AND NUMBER OF TESTEES IN EACH ABILITY GROUP

| Ability groups: Total score on SPM | White | Indian | Black |
|---|---|---|---|
| | N | N | N |
| 44-60 | 716 | 514 | 63 |
| 39-43 | 208 | 252 | 139 |
| 34-38 | 80 | 159 | 166 |
| 00-33 | 52 | 138 | 725 |
| Total | 1056 | 1063 | 1093 |

Because two population groups (white/Indian and white/black) and four ability categories are compared with each other on every item, the number of degrees of freedom are:

$df=(r-1)(k-1)$, where $r$ is the number of rows and $k$ is the number of columns.

$=(4-1)(2-1)$

$=3.$

The three-parameter logistical model

An explanation of the item response theory on which the three-parameter logistical model for identifying biased items is based can be found in Hambleton *et al.* (1978).

The mathematical form of the three-parameter logistical item characteristic curve is the following (Hambleton *et al.* 1978: 473):

$$P_g(\Theta) = c_g + (1-c_g)\frac{e^{Da_g(\Theta-b_g)}}{1+e^{Da_g(\Theta-b_g)}} \quad (g=1,2,\ldots,n).$$

where $P_g(\theta)$ = the probability that a testee with ability $\theta$ will answer item g correctly,

$D$ = a constant scaling factor (usually taken as 1,7),

$e$ = the mathematical constant 2,71828...,

$a_g$, $b_g$ and $c_g$ = the parameters that describe item g.

The a parameter is the discriminating ability of the item. The steeper the slope of the curve, the better the item can discriminate between good (large $\theta$) and poor (small $\theta$) testees.

The *b* parameter is an index of the degree of difficulty of the item and is represented by a point on the ability scale ($\theta$) that corresponds with the inflection point of the curve (the point at which the curve changes direction). The further the inflection point lies to the right, the more difficult the item, and the further it lies to the left, the easier the item. Although *b* usually lies between -3 and +3, the choice of this scale is arbitrary (Ree 1979).

The *c* parameter (pseudochance parameter) is the lower asymptote of the curve and specifies the probability that a testee with extremely limited ability can answer the item correctly by means of mere guesswork. This parameter is usually larger than 0,0 and smaller than 0,30.

In this investigation the item parameters ( *a*, *b* and *c* )as well as the ability parameter ($\theta$) were estimated with the aid of the computer program LOGIST (Wood, Wingersky & Lord 1976). These parameters are usually estimated simultaneously by means of a series of iterative steps. However the LOGIST program also allows one to make ability estimates ($\theta$) on the basis of a test other than the one involved. In the present study $\theta$ was based on the SPM, entered into the computer program separately, standardized ($\bar{x}=0,s=1$) and kept constant in all the subsequent phases of the estimating process. Because the $\theta$s for the *total* group were estimated on the basis of the SPM, the $\theta$s for the individual population groups were placed on the *same* scale. The *c* parameters (chance parameters) of the items of the Reasoning Test were estimated on the basis of the total group and kept constant. The *a* and *b* parameters of the items (in other words the discrimination and difficulty values) were estimated separately for each population group and compared with one another by means of the item characteristic curves (ICCs). The degree of bias in an item as reflected by the difference between the curves of two population groups (the white pupils served as reference group) was determined by means of a method suggested by Linn *et al.* (1980). Briefly, this method determines the area between the two ICCs by dividing the area between $\theta=-3$ and $\theta=+3$ into 600 intervals and calculating the following four indices:

1. Base Group Higher Area: the size of the area between $\theta=-3$ and $\theta=+3$ in which the ICC of the base group (in other words the reference group) lies higher than the ICC of the comparison group

2. Base Group Lower Area: the size of the area between $\theta=-3$ and $\theta=+3$ in which the ICC of the base group lies lower than the ICC of the comparison group

3. Absolute difference: the sum of Points 1 and 2 above

4. Square root of the Sum of Squares: the square root of the sum of the squared differences between the ICCs in the area $\theta=-3$ and $\theta=+3$.

An item with a large base group higher area (Index 1) but a small base group lower area (Index 2) is regarded as biased with respect to the comparison group. Such a finding indicates that there is a smaller probability of persons in the comparison group answering the item correctly than there is of persons in the base group with the same estimated ability ($\theta$) doing so. According to the above authors, items with a value of 0,2 or higher with respect to Index 4 are probably biased. The ICC values that appear in the tables in Section 5 of this report were based on Index 4.

As far as the interpretation of the reported indices of bias is concerned it should be kept in mind that the ability groups were formed on the basis of the SPM (which is a non-verbal test) and that the underlying trait that it measures will probably not be equally applicable to all the items in the Reasoning Test - particularly not in the case of verbal items.

## 5.0 RESULTS

The results are reported in four subsections. The first three of these concern potential or possible sources of item bias, namely the way in which an item is formulated, the item content and the item format. In the fourth subsection attention is given to the responses of testees to the item distractors.

## 5.1 THE WAY IN WHICH AN ITEM IS FORMULATED

This section reports results regarding how features of item formulation, such as the cues provided, the position of the correct answer among the distractors, etc., contribute to item bias.

### 5.1.1 Cues in test items

#### Hypothesis 1

According to Hypothesis 1 the various population groups do not benefit equally from the unintentional cues that are sometimes contained in an item. To investigate this hypothesis two pairs of items were used, namely Items 13 and 22 (first pair) and Items 32 and 42 (second pair). The two items of each pair were relatively similar, but the first item contained a cue that could help the testee in answering the question (cf. Appendix A1). In the case of the first pair the words "more than one" in Item 13 could cause the testees to select "double", and with respect to the second pair the prefix "pre" in "preconceived" (Item 32) could help the testee to select "prejudiced". Because the Afrikaans and the English versions of the two item pairs differ to a certain extent, the results of only the white English-speaking testees were compared with those of the other population groups. The results pertaining to the differences between the population groups in terms of the proportion of correct responses appear in Table 5.1a, and the bias indices ( $\chi^2_3$ and ICC) of the items involved in Table 5.1b. The percentage responses elicited by each distractor appear in Appendix B.

With regard to the first item pair, Table 5.1a reveals that all three population groups performed better in Item 13 (which contained a cue) than in Item 22. In the case of the white testees only 6 % answered Item 22 correctly, while 60 % answered Item 13 correctly. If the difference between the proportion of correct responses for the two items is

TABLE 5.1a

SIGNIFICANCE OF THE DIFFERENCES (Z) BETWEEN THE PROPORTIONS CORRECT (P) OF THE DIFFERENT POPULATION GROUPS WITH REGARD TO CUES IN TEST ITEMS

| Reasoning Test: Item no. | White (English) | | | Indian | | | Difference between White/Indian | Black | | | Difference between White/Black |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number correct | Number wrong | Proportion correct p | Number correct | Number wrong | Proportion correct p | $\left\lvert \begin{matrix} Z \\ p1-p2 \end{matrix} \right\rvert$ | Number correct | Number wrong | Proportion correct p | $\left\lvert \begin{matrix} Z \\ p1-p3 \end{matrix} \right\rvert$ |
| 13 1) | 321 | 212 | 0,60 | 294 | 769 | 0,27 | | 309 | 784 | 0,28 | |
| 22 | 32 | 501 | 0,06 | 69 | 994 | 0,06 | | 117 | 976 | 0,10 | |
| Difference in p | | | p1=0,54 | | | p2=0,21 | 13,37** | | | p3=0,18 | 14,87** |
| 32 1) | 123 | 410 | 0,23 | 163 | 900 | 0,15 | | 186 | 907 | 0,17 | |
| 42 | 197 | 336 | 0,37 | 309 | 754 | 0,29 | | 193 | 900 | 0,18 | |
| Difference in p | | | p1=-0,14 | | | p2=-0,14 | 0,0 | | | p3=-0,01 | 11,01** |

Note: 1) Item contains cue.
** $p<0,01$

TABLE 5.1b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.1a

| Item no. | Bias indices | | | |
|---|---|---|---|---|
| | $\chi^2_3$ | | ICC | |
| | Indian | Black | Indian | Black |
| 13 | 132,8* | 37,7* | 0,72+ | 0,73+ |
| 22 | 14,7* | 1,8 | 0,09 | 0,04 |
| 32 | 11,4* | 3,2 | 0,56+ | 0,55+ |
| 42 | 12,1* | 22,2* | 0,28+ | a |

Note: * $p<0,05$
+ICC>0,2
[a] Not possible to estimate item parameters.

studied, it is clear that the largest difference exists among the white testees, namely 0,54 ($p1$), followed by the Indian testees with 0,21 ($p2$) and the black testees with 0,18 ($p3$). The significance of the difference in proportions between white and Indian, and white and black testees was determined by means of the $z$ test for proportions for independent groups: $z = p1-p2/SD_p$, where $p1$ and $p2$ are the proportions of the two groups involved,

$p = N1p1+N2p2/N1+N2$ and

$$SD_p = \sqrt{p(1-p)\left(\frac{1}{N1} + \frac{1}{N2}\right)}$$

The significant $z$ values supported the conclusion that the white testees had utilized the cue in answering the item to a greater extent than the Indian and black testees. Therefore, Hypothesis 1 is confirmed with respect to the first item pair. In the case of Items 32 and 42 all three population groups actually found Item 42 (in other words the item without the cue) easier than Item 32 (which contained a cue). The difficulty of these two items apparently make them unsuitable for testing the above hypothesis. Hypothesis 1 was therefore confirmed only by one of the two item pairs, in which instance the white testees used the cue more effectively than the testees from the other population groups. The bias indices (Table 5.1b) reveal an interesting phenomenon, namely that the item containing a cue (Item 13) is apparently more biased than the item without the cue (Item 22). This can be attributed to the fact that the latter item is very difficult and was therefore answered correctly only by the best testees in each group. Group membership apparently did not play a role in the probability of answering the item correctly. According to Table 5.1a only 6 % of the white and Indian and 10 % of the black pupils answered Item 22 correctly.

### 5.1.2 Position of the correct answer among the distractors

#### Hypothesis 2

According to Hypothesis 2 the position of the correct answer among the distractors does not affect the performance of the various population groups in the same way.

In order to determine whether the position of the correct answer does affect the test performance of the various population groups four items of the JAT 2 (Reasoning) were repeated in the Reasoning Test. As a control measure the distractor positions were identical in two cases in the two tests, while they were changed in the other two cases: In Items 17 and 33 the correct answers were moved from the fourth to the first place and in Items 21 and 52 they were moved from the third to the fifth place. Because of a typing error in the Afrikaans version of Item 23 the results of only the English-speaking white testees for this item pair (Items 23 and 15) are given.

The results pertaining to the difficulty values of the items (p values) and the bias indices appear in Tables 5.2a and 5.2b respectively. Table 5.2a also contains the following information:

- *Odds ratio*. The odds ratio gives an indication of which member of the item pair has the greatest effect. The odds ratio is the ratio between the number of correct and the number of incorrect responses to the bottom member of the item pair, divided by the ratio between the number of correct and the number of incorrect responses to the top member of the item pair (Scheuneman 1985). If the manipulation of the item, in other words the changing of e.g. a concrete item to an abstract one, has no effect, the odds ratio is equal to one. A ratio of more than one indicates that the abstract (bottom) item is easier than the concrete (top) one, while a ratio of less than one indicates the opposite, in other words the concrete item is easier than the abstract one.

- The *p value* indicates the proportion of correct responses and is therefore the item's difficulty value.

- *Confidence interval for p*. The standard error of the sample ratio as an estimate of the population ratio is $\sqrt{p(1-p)/N}$. The maximum deviation that can appear in the present case (at the 5 % level of significance) is $1,96\sqrt{0,5(0,5)/1\ 000}$, in other words ± *0,031*. If the two intervals overlap, the two proportions concerned do not differ significantly from each other.

- *Chi-square*. In addition to the confidence interval the significance of the difference between two proportions was also calculated with the aid of a chi-square test for correlated proportions:
  $\chi^2 = (a-d)^2/a+d$, where a, b, c and d indicate the number of testees in a specific category, for example

| | | Item A | |
|---|---|---|---|
| | | Incorrect | Correct |
| Item B | Correct | a | b |
| | Incorrect | c | d |

From the results in Table 5.2a it is clear that with respect to identical items and answer positions the white and Indian testees tended to perform somewhat better with the second administration of the item (Reasoning Test) than with the first (JAT 2). In the case of the black pupils this tendency was reversed and they performed *less well* in the second than in the first administration. According to the confidence intervals the differences between the two administrations was significant only in one instance, and the above tendency could therefore be accidental.

TABLE 5.2a

ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR IDENTICAL ITEMS AND THE SAME/ DIFFERENT POSITIONS OF THE CORRECT ANSWERS

| (i) Identical items and positions of the correct answers | Item no. | White | | | | Indian | | | | Black | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ |
| Reasoning Test | 23 | 0,67 | 0,771 | 0,740-0,802[a] | 15,1** | 0,82 | 0,517 | 0,486-0,548[b] | 7,3** | 1,19 | 0,291 | 0,260-0,322[b] | 7,3** |
| JAT 2: Reasoning | 15 | | 0,694 | 0,663-0,725 | | | 0,475 | 0,444-0,506 | | | 0,340 | 0,309-0,371 | |
| Reasoning Test | 43 | 0,96 | 0,886 | 0,855-0,917[b] | 0,5 | 0,94 | 0,654 | 0,623-0,685[b] | 1,0 | 1,15 | 0,166 | 0,135-0,197[b] | 4,7* |
| JAT 2: Reasoning | 18 | | 0,881 | 0,850-0,912 | | | 0,645 | 0,614-0,676 | | | 0,196 | 0,165-0,227 | |
| (ii) Identical items but the positions of the correct answers differ | | | | | | | | | | | | | |
| Reasoning Test | 33 | 0,87 | 0,699 | 0,668-0,730[b] | 6,3* | 0,90 | 0,374 | 0,343-0,405[b] | 1,7 | 0,73 | 0,149 | 0,118-0,180[b] | 5,0* |
| JAT 2: Reasoning | 17 | | 0,668 | 0,637-0,699 | | | 0,354 | 0,323-0,385 | | | 0,117 | 0,086-0,148 | |
| Reasoning Test | 52 | 1,15 | 0,360 | 0,319-0,381[b] | 7,6* | 1,00 | 0,355 | 0,324-0,386[b] | 0,2 | 1,00 | 0,189 | 0,158-0,220[b] | 0,4 |
| JAT 2: Reasoning | 21 | | 0,388 | 0,357-0,419 | | | 0,356 | 0,325-0,387 | | | 0,189 | 0,158-0,220 | |

Note: [a] Differences between the two proportions significant ($p<0,05$).
[b] Differences between the two proportions not significant ($p>0,05$).
** $p<0,01$
* $p<0,05$

TABLE 5.2b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.2a

| Description | Item no. | Bias indices | | | |
|---|---|---|---|---|---|
| | | $\chi^2_3$ | | ICC | |
| | | Indian | Black | Indian | Black |
| (i) Identical items and positions of the correct answers | 23 | 12,3* | 4,69 | 0,28+ | 0,45+ |
| | 15 | 37,1* | 37,2* | 0,34+ | 0,78+ |
| | 43 | 31,2* | 253,8* | 0,54+ | 1,50+ |
| | 18 | 37,1* | 253,5* | 0,64+ | a |
| (ii) Identical items but the positions of the correct answers differ | 33 | 94,5* | 176,4* | 0,51+ | 1,24+ |
| | 17 | 89,4* | 187,7* | 0,54+ | 0,83+ |
| | 52 | 4,8 | 11,5* | 0,09 | 0,64+ |
| | 21 | 2,3 | 30,2* | 0,05 | a |

Note: *$p<0,05$
+ICC>0,2
[a] Not possible to estimate item parameters.

As far as identical items but different answer positions are concerned, the confidence intervals in Table 5.2a reveal that none of the differences between the p values of the first and second administrations of the item were significant. As the position of the correct answer is not at issue, Hypothesis 2 is not confirmed.

Table 5.2b reveals that bias is found in identical items as well as in items in which the position of the correct answer has been changed. Answer position as such is therefore not responsible for the bias, but the type of item used is (word classification was used in two of the cases and in the other two verbal analogies were used). Because all these items presuppose a certain knowledge of language, the bias could be due to a language deficit among the Indian and the black testees.

### 5.1.3 Attractiveness of item distractors

#### Hypothesis 3

According to Hypothesis 3 the population groups react differently with respect to the attractiveness of a specific distractor. To study this matter four items were selected at random from the Reasoning Test and a chi-square test, as described by Osterlind (1983: 70-73), was conducted for the number of testees in each population group who had selected a specific distractor. Owing to the large number of analyses involved it is not feasible to study the distractors of all the items; but the distractor responses given in Appendix B can be judged in the light of the 16 examples in Table 5.3.

TABLE 5.3

## SIGNIFICANCE OF THE DIFFERENCES ($\chi^2$) BETWEEN POPULATION GROUPS IN THE CHOICE OF ITEM DISTRACTORS

| Distractors | Item 4 | | | | Distractors | Item 8 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Indian | Black | $\chi^2$ | | White | Indian | Black | $\chi^2$ |
| A | 21 | 30 | 44 | 46,4 | A | 224 | 141 | 241 | 63,9 |
| B | 28 | 84 | 246 | 412,9 | B | 108 | 126 | 102 | 3,5[a] |
| C | 27 | 99 | 182 | 272,1 | C1) | 671 | 737 | 485 | - |
| D | 42 | 129 | 233 | 310,5 | D | 8 | 6 | 43 | 69,7 |
| E1) | 937 | 719 | 382 | - | E | 45 | 52 | 220 | 239,6 |
| **Distractors** | **Item 20** | | | | **Distractors** | **Item 66** | | | |
| | White | Indian | Black | $\chi^2$ | | White | Indian | Black | $\chi^2$ |
| A | 114 | 141 | 167 | 62,2 | A1) | 328 | 209 | 88 | - |
| B1) | 561 | 459 | 278 | - | B | 239 | 169 | 140 | 25,1 |
| C | 115 | 124 | 223 | 123,6 | C | 227 | 208 | 175 | 46,9 |
| D | 154 | 183 | 194 | 52,9 | D | 45 | 145 | 358 | 386,4 |
| E | 106 | 146 | 217 | 116,6 | E | 212 | 328 | 308 | 143,0 |

Note: 1) Correct answer
[a] Not significant; all other values significant ($p<0,01$).

The results in Table 5.3 clearly indicate that with one exception, namely Distractor B of Item 8, the three population groups differed significantly from one another with regard to their choice of distractors. Hypothesis 3 is therefore confirmed. It is also clear that the significant result was caused mainly by the responses of the black testees. These testees tended more often than the white and the Indian testees to select less possible answers such as Distractors B and D of Item 4, Distractor E of Item 8, Distractors C and E of Item 20 and Distractors D and E of Item 66.

### 5.1.4 True and false answers

#### Hypothesis 4

According to this hypothesis the various population groups do not react in the same manner to items requiring the identification of an incorrect (false) answer rather than a correct (true) one.

The results in Table 5.4a do not confirm this hypothesis. All the testees, irrespective of their group membership, found the item requiring the identification of the true answer considerably easier than a similar item requiring the identification of the false answer. For example: 69 % of the white testees answered Item 3 correctly, while only 46 % of these testees succeeded with Item 11. Of the Indian testees the corresponding percentages were 53 % and 39 % and of the black testees 30 % and 23 %. This trend was confirmed by the second item pair (Items 30 and 40) with regard to the white and the Indian testees. The fact that the same trend did not apply to the black testees should be attributed to their finding both items extremely difficult (only 20 % answered them correctly) and it made no difference to their performance whether the true or false answer had to be identified. Therefore the results in Table 5.4a draw attention to the special problems black testees experience with items that involve only language skills (cf. Appendix A2).

In the case of the Indian testees the ICC index (Table 5.4b) reveals that all the items are unbiased. However the $X_3^2$ index indicates significant values, and contrary to general expectation these are for items that require identification of the one true answer. In the case of the black testees significant values were found for all four of the items, but both the $X_3^2$ and the ICC indices indicated higher values with respect to the true answer type than the false answer type. This unexpected result can perhaps be attributed to the fact that black testees find both item types very difficult and that the bias indices are therefore relatively unstable (the differences between the two indices could be quite coincidental). The bias of the items can be attributed to the influence of the language factor.

TABLE 5.4a

ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$ )WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR TRUE/FALSE ANSWER OPTIONS

| Reasoning Test: True/false answers | Item no. | White | | | | Indian | | | | Black | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ |
| One true answer | 3 | 0,39 | 0,689 | 0,658-0,720[a] | 141,6** | 0,58 | 0,531 | 0,500-0,562[a] | 50,0** | 0,70 | 0,304 | 0,273-0,335[a] | 14,5** |
| One false answer | 11 | | 0,458 | 0,427-0,489 | | | 0,394 | 0,363-0,425 | | | 0,232 | 0,201-0,263 | |
| One true answer | 30 | 0,32 | 0,768 | 0,737-0,799[a] | 170,2** | 0,61 | 0,573 | 0,542-0,604[a] | 44,1** | 1,05 | 0,197 | 0,166-0,228[b] | 0,2 |
| One false answer | 40 | | 0,513 | 0,482-0,544 | | | 0,449 | 0,418-0,480 | | | 0,205 | 0,174-0,236 | |
| Mean proportion -True: | | | 0,729 | 0,698-0,760[a] | | | 0,552 | 0,521-0,583[a] | | | 0,251 | 0,220-0,282[b] | |
| -False: | | | 0,486 | 0,455-0,517 | | | 0,422 | 0,391-0,453 | | | 0,219 | 0,188-0,250 | |

Note: [a] Differences between the two proportions significant (p<0,05).
[b] Differences between the two proportions not significant (p>0,05).
** p<0,01
* p<0,05

TABLE 5.4b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.4a

| Description | Item no. | Bias indices | | | |
|---|---|---|---|---|---|
| | | $\chi^2_3$ | | ICC | |
| | | Indian | Black | Indian | Black |
| One true answer | 3 | 13,8* | 30,9* | 0,17 | 0,87+ |
| One false answer | 11 | 4,7 | 19,1* | 0,08 | a |
| One true answer | 30 | 17,6* | 93,9* | 0,19 | 0,58+ |
| One false answer | 40 | 0,6 | 20,0* | 0,10 | 0,69+ |

Note: *p<0,05
+ICC>0,2
[a] Not possible to estimate item parameters.

### 5.1.5 Positive and negative items

#### Hypothesis 5

According to Hypothesis 5 the population groups differ in their reactions to positively and negatively phrased items.

The results appear in Tables 5.5a and 5.5b and the test items appear in Appendix A3. It is clear from Table 5.5a that the differences between the proportions of positive and negative items were consistently significant among the white testees, but not among the Indian and the black testees. Since the same trend was found for all three population groups (namely, that positively phrased items are generally easier than similar items negatively phrased), it can be concluded that the groups react largely in the same way, and Hypothesis 5 is not confirmed.

An interesting result appears in Table 5.5b, namely that for the Indian testees all the positive items were biased and almost all the negative items were unbiased according to the $\chi^2_3$ index. The ICC index displayed a measure of bias for two positive items and one negative item. It appears that although positive items were the easier of the two types, Indian testees were not as likely as white testees to answer them correctly. Their chances of answering the more difficult (that is, negative) items correctly were however more or less the same as those of the white testees. Black testees, on the other hand, were less likely than whites to answer correctly, regardless of item type - probably due to a language deficit.

TABLE 5.5a

ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR POSITIVE/NEGATIVE TEST ITEMS

| Reasoning Test: Positive/negative items | Item no. | White |  |  |  | Indian |  |  |  | Black |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ |
| Positive | 3 | 0,53 | 0,689 | 0,658-0,720$^a$ | 66,3** | 0,69 | 0,531 | 0,500-0,562$^a$ | 23,9** | 0,80 | 0,304 | 0,273-0,335$^b$ | 6,5* |
| Negative | 20 |  | 0,534 | 0,503-0,565 |  |  | 0,436 | 0,405-0,467 |  |  | 0,258 | 0,227-0,289 |  |
| Positive | 5 | 0,73 | 0,648 | 0,617-0,679$^a$ | 14,7** | 0,95 | 0,473 | 0,442-0,504$^b$ | 0,4 | 0,48 | 0,339 | 0,308-0,370$^a$ | 59,6** |
| Negative | 44 |  | 0,572 | 0,541-0,603 |  |  | 0,460 | 0,429-0,491 |  |  | 0,198 | 0,167-0,229 |  |
| Positive | 30 | 0,48 | 0,768 | 0,737-0,798$^a$ | 72,1** | 0,77 | 0,573 | 0,542-0,604$^a$ | 10,7** | 1,52 | 0,197 | 0,166-0,228$^a$ | 18,4** |
| Negative | 50 |  | 0,615 | 0,584-0,646 |  |  | 0,508 | 0,477-0,539 |  |  | 0,274 | 0,241-0,305 |  |
| Positive | 59 | 0,72 | 0,479 | 0,438-0,510$^a$ | 19,2** | 1,10 | 0,296 | 0,265-0,327$^b$ | 0,9 | 0,76 | 0,304 | 0,273-0,335$^b$ | 8,1** |
| Negative | 67 |  | 0,396 | 0,365-0,427 |  |  | 0,314 | 0,283-0,345 |  |  | 0,250 | 0,219-0,281 |  |
| Mean proportion - Positive: |  |  | 0,646 | 0,615-0,677$^a$ |  |  | 0,468 | 0,437-0,499$^b$ |  |  | 0,286 | 0,255-0,317$^b$ |  |
| - Negative: |  |  | 0,529 | 0,498-0,560 |  |  | 0,430 | 0,399-0,461 |  |  | 0,245 | 0,214-0,276 |  |

Note: $^a$Differences between the two proportions significant ($p<0,05$).
$^b$Differences between the two proportions not significant ($p>0,05$).
** $p<0,01$
* $p<0,05$

TABLE 5.5b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.5a

| Description | Item no. | Bias indices |  |  |  |
|---|---|---|---|---|---|
|  |  | $\chi^2_3$ |  | ICC |  |
|  |  | Indian | Black | Indian | Black |
| Positive | 3 | 13,8* | 30,9* | 0,17 | 0,87+ |
| Negative | 20 | 1,9 | 8,0* | 0,05 | 0,67+ |
| Positive | 5 | 18,5* | 16,4* | 0,23+ | 0,53+ |
| Negative | 44 | 8,6* | 44,2* | 0,13 | 0,34+ |
| Positive | 30 | 17,6* | 93,9* | 0,19 | 0,58+ |
| Negative | 50 | 3,5 | 22,2* | 0,08 | 0,83+ |
| Positive | 59 | 30,5* | 8,5* | 0,27+ | 0,72+ |
| Negative | 67 | 6,0 | 3,8 | 0,32+ | 0,53+ |

Note: *$p<0,05$
+ICC>0,2

## 5.2 ITEM CONTENT

This section reports the effects on item bias of items with the same format (for example, syllogistic) but different contents (for example, concrete versus abstract).

### 5.2.1 Concrete and abstract items

Hypothesis 6
According to Hypothesis 6(a) concrete and abstract items do not have the same effect on test performance, and (b) population groups differ from one another in their reactions to concrete and abstract items.
The possibility that the test performance of the three population groups is affected differently by concrete and abstract item content was studied with respect to syllogisms. The results are given in Tables 5.6a and 5.6b and the test items appear in Appendix A4. It is clear from Table 5.6a that both the white and the black testees displayed significant differences with respect to only two item pairs (both groups found the concrete item easier in the one case and the abstract item easier in the other), while the Indian testees displayed significant differences with respect to four item pairs (in two cases the concrete item and in two cases the abstract item was easier). None of the population groups shared a significant difference between the mean proportion for concrete and abstract items. Because concrete and abstract items apparently have the same effect on test performance, Hypothesis 6(a) is not confirmed. A graphical representation of the mean item difficulty values of concrete and abstract syllogisms (Figure 5.1) also reveals that the population groups did not differ in their reactions to concrete and abstract items, and Hypothesis 6(b) is therefore not confirmed either.

TABLE 5.6a

ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$ ) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR CONCRETE AND ABSTRACT SYLLOGISMS

| Reasoning Test: Concrete and Abstract Syllogisms | Item no. | White | | | | Indian | | | | Black | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ |
| Concrete<br>Abstract | 5<br>14 | 1,48 | 0,648<br>0,732 | 0,617-0,679[a]<br>0,701-0,763 | 21,90** | 1,48 | 0,473<br>0,571 | 0,442-0,504[a]<br>0,540-0,602 | 26,72** | 0,76 | 0,339<br>0,282 | 0,308-0,370[b]<br>0,251-0,313 | 10,25** |
| Concrete<br>Abstract | 24<br>34 | 0,95 | 0,318<br>0,310 | 0,287-0,349[b]<br>0,279-0,341 | 0,27 | 0,75 | 0,244<br>0,196 | 0,213-0,275[b]<br>0,165-0,227 | 8,86** | 1,65 | 0,145<br>0,219 | 0,114-0,176[a]<br>0,188-0,250 | 20,59** |
| Concrete<br>Abstract | 44<br>53 | 0,83 | 0,572<br>0,527 | 0,541-0,603[b]<br>0,496-0,558 | 6,63** | 0,81 | 0,460<br>0,408 | 0,429-0,491[b]<br>0,377-0,439 | 7,08** | 1,22 | 0,198<br>0,234 | 0,167-0,229[b]<br>0,203-0,265 | 5,09* |
| Concrete<br>Abstract | 61<br>68 | 0,94 | 0,506<br>0,494 | 0,475-0,537[b]<br>0,463-0,525 | 0,52 | 1,36 | 0,312<br>0,383 | 0,281-0,343[a]<br>0,352-0,414 | 19,56** | 1,52 | 0,133<br>0,191 | 0,102-0,164[b]<br>0,160-0,222 | 15,37** |
| Concrete<br>Abstract | 74<br>79 | 0,61 | 0,640<br>0,520 | 0,609-0,671[a]<br>0,489-0,551 | 47,75** | 0,61 | 0,505<br>0,385 | 0,474-0,536[a]<br>0,354-0,416 | 43,57** | 0,91 | 0,226<br>0,211 | 0,195-0,257[b]<br>0,180-0,242 | 1,08 |
| Concrete<br>Abstract | 81<br>83 | 0,84 | 0,370<br>0,332 | 0,339-0,401[b]<br>0,301-0,363 | 4,92* | 0,63 | 0,364<br>0,265 | 0,333-0,395[a]<br>0,234-0,296 | 31,23** | 0,53 | 0,220<br>0,132 | 0,189-0,251[a]<br>0,101-0,163 | 37,44** |
| Mean proportion - Concrete:<br>- Abstract: | | | 0,509<br>0,486 | 0,478-0,540[b]<br>0,455-0,517 | | | 0,393<br>0,368 | 0,362-0,424[b]<br>0,337-0,399 | | | 0,210<br>0,211 | 0,179-0,241[b]<br>0,180-0,242 | |

Note: [a] Differences between the two proportions significant ($p<0,05$).
[b] Differences between the two proportions not significant ($p>0,05$).
** $p<0,01$
* $p<0,05$

TABLE 5.6b

BIAS INDICES ($X^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.6a

| Description | Item no. | Bias indices | | | |
|---|---|---|---|---|---|
| | | $X^2_3$ | | ICC | |
| | | Indian | Black | Indian | Black |
| Concrete | 5 | 18,5* | 16,4* | 0,23+ | 0,53+ |
| Abstract | 14 | 11,5* | 31,1* | 0,18 | 0,27+ |
| Concrete | 24 | 4,7 | 9,6* | 0,08 | 0,17 |
| Abstract | 34 | 19,3* | 4,6 | 0,34+ | 0,53+ |
| Concrete | 44 | 8,6* | 44,2* | 0,13 | 0,34+ |
| Abstract | 53 | 9,5* | 17,4* | 0,07 | 0,78+ |
| Concrete | 61 | 20,5* | 29,7* | 0,14 | 0,40+ |
| Abstract | 68 | 7,3 | 23,1* | 0,06 | 0,52+ |
| Concrete | 74 | 6,8 | 38,5* | 0,15 | 0,37+ |
| Abstract | 79 | 15,2* | 32,9* | 0,13 | 0,41+ |
| Concrete | 81 | 0,6 | 22,8* | 0,25+ | a |
| Abstract | 83 | 6,4 | 25,6* | 0,28+ | 0,48+ |

Note: $^*p<0,05$
$^+ICC>0,2$
[a] Not possible to estimate item parameters.

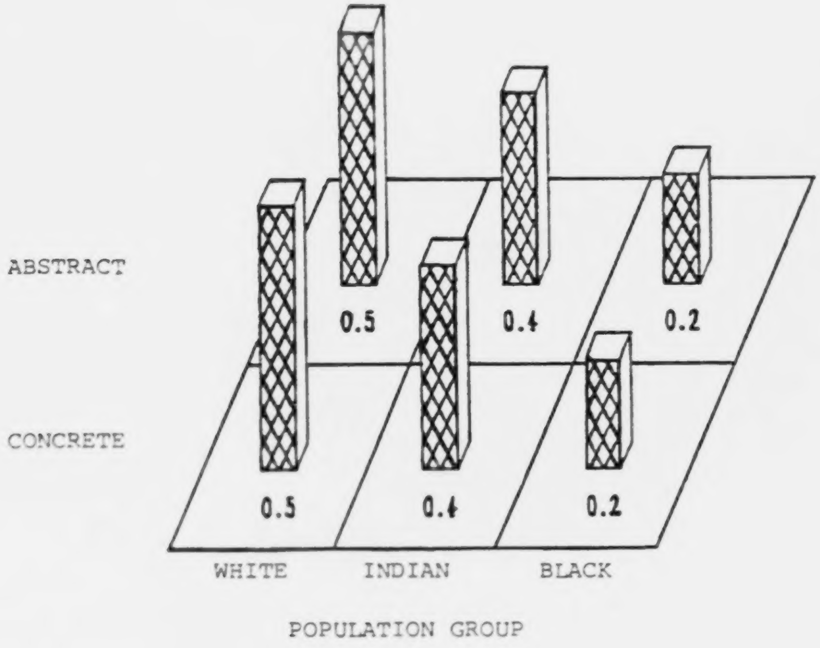


FIGURE 5.1: GRAPHICAL REPRESENTATION OF THE MEAN ITEM DIFFICULTY VALUES FOR CONCRETE AND ABSTRACT SYLLOGISMS

The results with regard to item bias appear in Table 5.6b. In the case of the Indian testees the ICC index identified only four of the 12 items as possibly biased. Although all the content types were generally suitable for these testees, it nevertheless appears that concrete-meaningless, concrete-unrealistic and abstract contents are somewhat less suitable than, for example, concrete-meaningful material.

The results indicate that syllogisms are generally not a suitable item format for measuring reasoning ability in black testees. With the exception of Item 24 all the items have significant bias indices(ICC). The content of this item, concrete meaningful, thus appears to be the only content type suitable for black testees.

Hypothesis 7

The results in Table 5.6a also do not serve to confirm Hypothesis 7, according to which black testees perform relatively better in concrete than in abstract items. Not only the black but also the white and Indian testees find the abstract items just as difficult as the concrete ones.

### 5.2.2 The use of words and letters in test items as opposed to the use of numbers and figures

Hypothesis 8

According to Hypothesis 8 the performance of different population groups is differently affected by verbal and non-verbal items. This hypothesis was studied with the aid of letter as opposed to number series and verbal as opposed to figure analogies. Two item formats were therefore used, namely series and analogies. The contents varied within these formats, namely verbal(words) and non-verbal (numbers and figures), corresponded closely enough to require the same item solution strategy (cf. test items in Appendix A5). Any differential effect can therefore be attributed to the manipulation of the item content.

As far as letter and figure series are concerned, Table 5.7a reveals that the confidence intervals for each of the item pairs overlap, indicating no significant differences (the significant $\chi^2$ values that appear in places can be ignored because of the overlap). It can be concluded that it makes no difference to the testees' performance whether a series is presented in the form of letters or numbers.

Another notable feature of Table 5.7a is that the black testees found series as an item format relatively easy - as can be seen from the proportion of correct responses, which varies between 0,5 and 0,8. The only exception here was the item pair 65 and 57. However this item pair is more complex than the others and white and Indian testees also experienced problems with it.

As regards verbal analogies as opposed to figure analogies (Table 5.7a), it is clear from the odds ratio and the confidence intervals that the manipulation of item content had a significant effect on test performance in some cases. This effect was identified with respect to the following items and population groups:

TABLE 5.7a

ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$ ) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR LETTER AND NUMBER SERIES, VERBAL AND NON-VERBAL (FIGURE) ANALOGIES

| (i) Letter-Number Series | Type | Item no. | White | | | | Indian | | | | Black | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ |
| Reasoning Test RT | Letter | 28 | 0,85 | 0,946 | 0,915-0,977[b] | 0,9 | 3,46 | 0,914 | 0,883-0,945[b] | 35,6** | 0,91 | 0,804 | 0,773-0,835[b] | 2,1 |
| JAT 2: Reasoning | Number | 2 | | 0,936 | 0,905-0,967 | | | 0,973 | 0,942-0,999 | | | 0,781 | 0,750-0,812 | |
| RT | Letter | 10 | 0,61 | 0,895 | 0,864-0,926[b] | 25,3** | 0,77 | 0,841 | 0,810-0,872[b] | 11,7** | 0,20 | 0,551 | 0,520-0,582[b] | 5,9* |
| RT | Number | 2 | | 0,836 | 0,805-0,867 | | | 0,800 | 0,769-0,831 | | | 0,516 | 0,485-0,547 | |
| RT | Letter | 48 | 1,16 | 0,909 | 0,878-0,940[b] | 1,2 | 1,09 | 0,814 | 0,783-0,845[b] | 1,1 | 0,91 | 0,549 | 0,518-0,580[b] | 2,9 |
| RT | Number | 38 | | 0,920 | 0,889-0,951 | | | 0,827 | 0,796-0,858 | | | 0,522 | 0,491-0,553 | |
| RT | Letter | 65 | 1,22 | 0,260 | 0,229-0,291[b] | 6,6** | 1,08 | 0,262 | 0,231-0,293[b] | 1,3 | 1,04 | 0,136 | 0,105-0,167[b] | 0,1 |
| RT | Number | 57 | | 0,298 | 0,267-0,329 | | | 0,279 | 0,248-0,310 | | | 0,141 | 0,110-0,172 | |
| Mean proportion | -Letter: | | | 0,753 | 0,722-0,784[b] | | | 0,708 | 0,677-0,739[b] | | | 0,510 | 0,479-0,541[b] | |
| | -Number: | | | 0,748 | 0,717-0,779 | | | 0,720 | 0,689-0,751 | | | 0,490 | 0,459-0,521 | |
| (ii) Reasoning Test Verbal-Figure Analogies | | | | | | | | | | | | | | |
| | V | 6 | 0,83 | 0,526 | 0,495-0,557[b] | 5,6* | 0,89 | 0,364 | 0,333-0,395[b] | 2,1 | 1,33 | 0,184 | 0,153-0,215[b] | 9,1** |
| | F | 15 | | 0,481 | 0,450-0,512 | | | 0,336 | 0,305-0,397 | | | 0,233 | 0,202-0,264 | |
| | V | 25 | 3,02 | 0,446 | 0,415-0,477[a] | 196,2** | 1,01 | 0,692 | 0,661-0,723[b] | 0,1 | 0,65 | 0,501 | 0,470-0,532[a] | 26,4** |
| | F | 35 | | 0,711 | 0,680-0,742 | | | 0,697 | 0,666-0,728 | | | 0,397 | 0,365-0,428 | |
| | V | 27 | 0,89 | 0,728 | 0,697-0,759[b] | 1) | 0,94 | 0,596 | 0,565-0,627[b] | 1) | 0,97 | 0,311 | 0,280-0,342[b] | 1) |
| | F | 37 | | 0,703 | 0,672-0,734 | | | 0,582 | 0,551-0,613 | | | 0,301 | 0,270-0,332 | |
| | V | 45 | 0,80 | 0,567 | 0,536-0,598[b] | 10,8** | 0,54 | 0,549 | 0,518-0,580[a] | 76,4** | 0,50 | 0,401 | 0,370-0,432[a] | 56,5** |
| | F | 54 | | 0,516 | 0,485-0,547 | | | 0,400 | 0,369-0,431 | | | 0,255 | 0,224-0,286 | |
| | V | 62 | 1,79 | 0,252 | 0,221-0,283[a] | 61,2** | 1,65 | 0,156 | 0,125-0,187[a] | 26,9** | 0,99 | 0,138 | 0,107-0,169[b] | 0,0 |
| | F | 69 | | 0,377 | 0,346-0,408 | | | 0,233 | 0,202-0,264 | | | 0,138 | 0,107-0,169 | |
| | V | 75 | 1,26 | 0,658 | 0,627-0,689[b] | 8,1** | 2,46 | 0,371 | 0,340-0,402[a] | 146,5** | 2,93 | 0,107 | 0,076-0,138[a] | 97,0** |
| | F | 82 | | 0,708 | 0,677-0,739 | | | 0,594 | 0,563-0,625 | | | 0,265 | 0,234-0,296 | |
| Mean proportion | - Verbal: | | | 0,529 | 0,498-0,560[b] | | | 0,455 | 0,424-0,486[b] | | | 0,274 | 0,243-0,305[b] | |
| | - Figure: | | | 0,583 | 0,552-0,614 | | | 0,474 | 0,443-0,505 | | | 0,265 | 0,234-0,296 | |

Note: [a] Differences between the two proportions significant ($p<0,05$).

[b] Differences between the two proportions not significant ($p>0,05$).

** $p<0,01$

* $p<0,05$

1) $\chi^2$ not calculated.

- Items 25 and 35 for white and black pupils. However the sequence of the item difficulty values differed for the two groups. For white testees the figure analogy was the easiest and for black testees the verbal analogy.

- Items 45 and 54 for Indian and black testees (the verbal analogy was the easier item for both groups).

- Items 62 and 69 for white and Indian testees (the figure analogy was the easier item for both groups).

- Items 75 and 82 for Indian and black testees (in both cases the figure analogy was the easier item).

It seems that in three of the six cases, white testees found the figure analogies easier than the corresponding verbal analogies. Although the chi-square test indicated significant values in five cases, there were only two differences worth mentioning according to the confidence intervals, namely between Items 25 and 35 and between Items 62 and 69. Three significant differences were found for the Indian testees: in two cases the figure analogy was easier than the verbal analogy.

The black testees' performance deviated to a certain extent from that of the other two groups: although there were also three significant differences, the figure analogy was easier than the verbal analogy in only one case (Item 82).

It is clear from the above that the verbal and the figure analogies did not affect the performance of the different population groups in exactly the same way and Hypothesis 8 is therefore confirmed to a certain extent.

Numerous analyses carried out so far have revealed a language deficit among black testees. They were thus expected to do better in figure than in verbal analogies. The fact that the results in Table 5.7a do not support this expectation should probably be attributed to the degree of difficulty of the items concerned. In the case of the white testees eight of the twelve proportions were greater than 0,5; only one such proportion was found for the black testees. In contrast, seven of the proportions for the black testees were less than 0,3, while this figure occured only once for the white testees. From this it is evident that the mere replacement of words by figures in an analogy item will not increase the percentage of black testees answering the item correctly. Results pertaining to bias appear in Table 5.7b.

TABLE 5.7b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.7a

| Description | Item no. | Bias indices | | | |
|---|---|---|---|---|---|
| | | $\chi^2_3$ | | ICC | |
| | | Indian | Black | Indian | Black |
| (i) Letter-Number Series | | | | | |
| Letter | 28 | 0,5 | 2,3 | 0,10 | 0,18 |
| Number (JAT 2) | 2 | 1,5 | 1,7 | 0,43+ | 0,21 |
| Letter | 10 | 0,7 | 4,7 | 0,07 | 0,19 |
| Number | 2 | 0,6 | 4,8 | 0,06 | 0,12 |
| Letter | 48 | 2,9 | 12,9* | 0,11 | 0,33+ |
| Number | 38 | 2,8 | 17,1* | 0,16 | 0,37+ |
| Letter | 65 | 5,4 | 2,4 | 0,14 | 0,37+ |
| Number | 57 | 1,3 | 5,9 | 0,10 | 0,26+ |
| (ii) Verbal-Figure Analogies | | | | | |
| Verbal | 6 | 31,7* | 83,3* | 0,33+ | a |
| Figure | 15 | 16,7* | 11,9* | 0,19 | 0,28+ |
| Verbal | 25 | 64,8* | 7,3 | 0,57+ | 0,35+ |
| Figure | 35 | 5,3 | 14,1* | 0,47+ | 0,45+ |
| Verbal | 27 | 5,9 | 22,3* | 0,15 | 0,31+ |
| Figure | 37 | 5,7 | 23,8* | 0,15 | 0,36+ |
| Verbal | 45 | 3,0 | 4,4 | 0,10 | 0,47+ |
| Figure | 54 | 6,0 | 5,4 | 0,10 | 0,12 |
| Verbal | 62 | 15,0* | 17,3* | 0,11 | a |
| Figure | 69 | 28,3* | 20,3* | 0,38+ | a |
| Verbal | 75 | 80,8* | 148,7* | 0,34+ | 0,49+ |
| Figure | 82 | 2,0 | 26,1* | 0,09 | 0,36+ |

Note: *p<0,05
+ICC>0,2
aNot possible to estimate item parameters.

As far as letter and number series are concerned few biased items were identified and the indices are generally quite small. Series, regardless of the content type, therefore appears to have been a reasonably suitable item format for most of the Indian and black testees. However the same generalization cannot be made with regard to analogies. In the case of the Indian testees verbal analogies showed a greater incidence of bias than figure analogies, while for black testees approximately the same number of verbal and figure analogies were biased. This finding corresponds with that of Scheuneman (1980, 1981) among others, and illustrates the point made by Bond (1981), namely that it is impossible to separate vocabulary from verbal-analogous reasoning, although the two differ conceptually.

### 5.2.3 Concepts with regard to time

#### Hypothesis 9

According to Hypothesis 9 black testees perform worse than white testees in test items in which time concepts play a role. The results appear in Tables 5.8a and 5.8b and the test items are included in Appendix A6.

The mean proportions of time and "other" concepts (Table 5.8a) reveal that none of the three groups found time concepts significantly more difficult than the "other" concepts. If the mean proportion for time concepts is expressed as a proportion of time plus other concepts, it is also clear that time concepts are not relatively more difficult for black testees than for white or Indian testees
(white: 0,544/(0,544 + 0,599) =0,48,
Indian: 0,461/(0,461 + 0,493) =0,48,
black: 0,292/(0,292 + 0,257) =0,53).
The black testees therefore found time concepts relatively easier than the white testees, and Hypothesis 9 is not confirmed. Although black testees achieve somewhat better scores in items involving time concepts rather than "other" concepts, it can be concluded from the size of the two mean proportions that they do experience problems with both item types (29 % and 26 % of the testees respectively answered these two item types correctly).

Certain item pairs revealed interesting trends:

- Approximately 36 % of the black testees answered Item 26 correctly, as opposed to 47 % who answered Item 36 correctly. Since exactly the same numerical calculations have to be performed for both items and the only differences between them are that the one involves the number of hours and minutes and the other the number of litres, it is clear that black testees are not yet very familiar with the new way of expressing time (for instance 06h15). The same trend was identified among the Indian and white testees, but because they found the items easier than the black testees, the effect of the new method is less noticeable.

TABLE 5.8a

ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^{2}$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR TIME CONCEPTS

| Reasoning Test: Time Concepts in different Modes | Item no. | White | | | | Indian | | | | Black | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Odds ratio | p | 95 % confidence interval for p | $\chi^{2}$ | Odds ratio | p | 95 % confidence interval for p | $\chi^{2}$ | Odds ratio | p | 95 % confidence interval for p | $\chi^{2}$ |
| (i) Verbal-Number Mode | | | | | | | | | | | | | |
| Arithmetic problem (time in words) | 70 | 1,3 | 0,331 | 0,300-0,362$^{b}$ | 12,6** | 0,89 | 0,247 | 0,216-0,278$^{b}$ | 1,8 | 1,3 | 0,157 | 0,126-0,188$^{b}$ | 4,0* |
| Arithmetic problem (time in numbers) | 63 | | 0,392 | 0,361-0,423 | | | 0,225 | 0,194-0,256 | | | 0,194 | 0,163-0,225 | |
| (ii) Verbal-Figure Mode | | | | | | | | | | | | | |
| Verbal analogy (time in words) | 27 | 0,89 | 0,728 | 0,697-0,759$^{b}$ | 2,2 | 0,94 | 0,598 | 0,567-0,629$^{b}$ | 0,5 | 0,97 | 0,311 | 0,280-0,342$^{b}$ | 0,0 |
| Figure analogy (time:-figure) | 37 | | 0,704 | 0,673-0,735 | | | 0,585 | 0,554-0,616 | | | 0,307 | 0,276-0,338 | |
| (iii) Time concept - Other concept | | | | | | | | | | | | | |
| Classification (time concept) | 8 | 1,71 | 0,635 | 0,604-0,666$^{a}$ | 32,7** | 0,80 | 0,694 | 0,663-0,725$^{b}$ | 6,3* | 0,61 | 0,445 | 0,414-0,476$^{a}$ | 33,3** |
| Classification (other concept - names) | 17 | | 0,750 | 0,719-0,781 | | | 0,644 | 0,613-0,675 | | | 0,328 | 0,297-0,359 | |
| Number series (time in numbers) | 26 | 1,36 | 0,764 | 0,733-0,795$^{b}$ | 10,2** | 1,50 | 0,669 | 0,638-0,700$^{a}$ | 25,5** | 1,56 | 0,365 | 0,334-0,396$^{a}$ | 39,0** |
| Number series (litres) | 36 | | 0,815 | 0,784-0,846 | | | 0,754 | 0,723-0,785 | | | 0,472 | 0,431-0,503 | |
| Arithmetic problem (time in numbers) | 46 | 2,12 | 0,404 | 0,373-0,435$^{a}$ | 89,7** | 1,71 | 0,319 | 0,288-0,350$^{a}$ | 42,0** | 0,59 | 0,231 | 0,200-0,262$^{a}$ | 20,2** |
| Arithmetic problem (money) | 55 | | 0,592 | 0,561-0,623 | | | 0,444 | 0,413-0,475 | | | 0,154 | 0,123-0,185 | |
| Verbal analogy (days of week) | 76 | 0,52 | 0,375 | 0,344-0,406$^{a}$ | 67,6** | 0,77 | 0,163 | 0,132-0,194$^{b}$ | 6,4* | 0,55 | 0,128 | 0,097-0,159$^{b}$ | 17,4** |
| Verbal analogy (school standards) | 80 | | 0,239 | 0,208-0,270 | | | 0,130 | 0,099-0,161 | | | 0,074 | 0,043-0,105 | |
| Mean proportion - Time Concept: | | | 0,544 | 0,513-0,575$^{b}$ | | | 0,461 | 0,430-0,492$^{b}$ | | | 0,292 | 0,261-0,323$^{b}$ | |
| - Other Concept: | | | 0,599 | 0,568-0,630 | | | 0,493 | 0,462-0,524 | | | 0,257 | 0,226-0,288 | |

Note: [a] Differences between the two proportions significant ($p<0,05$).
[b] Differences between the two proportions not significant ($p>0,05$)
** $p<0,01$
* $p<0,05$

TABLE 5.8b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.8a

| Description | Item no. | Bias indices | | | |
|---|---|---|---|---|---|
| | | $\chi^2_3$ | | ICC | |
| | | Indian | Black | Indian | Black |
| (i) Verbal - Number Mode | | | | | |
| Verbal | 70 | 9,5* | 13,3* | 0,06 | 0,15 |
| Number | 63 | 31,1* | 14,6* | 0,15 | 0,86+ |
| (ii) Verbal - Figure Mode | | | | | |
| Verbal | 27 | 5,9 | 22,3* | 0,15 | 0,31+ |
| Figure | 37 | 5,7 | 23,8* | 0,15 | 0,36+ |
| (iii) Time concept-Other concept | | | | | |
| Time | 8 | 5,7 | 6,7 | 0,35+ | 0,23+ |
| Other | 17 | 8,3* | 68,1* | 0,36+ | 0,80+ |
| Time | 26 | 4,2 | 21,9* | 0,21+ | 0,30+ |
| Other | 36 | 0,5 | 7,5 | 0,04 | 0,27+ |
| Time | 46 | 8,1* | 9,4* | 0,07 | 0,75+ |
| Other | 55 | 9,0* | 56,1* | 0,11 | 0,38+ |
| Time | 76 | 65,9* | 24,5* | 0,25+ | 0,54+ |
| Other | 80 | 20,8* | 11,7* | 0,20 | 0,39+ |

Note: $^*p<0,05$
+ICC>0,2

- According to the results of Items 27 and 37 the performance of all three population groups is uneffected by whether the time concept concerned is presented as a verbal or as a figure analogy.

Table 5.8b clearly indicates that few items with large bias indices are found in the case of the Indian testees: both the $\chi_3^2$ and ICC indices identify only three such items. In the case of the black testees most of the items were probably biased. However in view of the degree of difficulty of these items no special value can be attached to these indices of bias.

### 5.2.4 Direction and orientation

Hypothesis 10

In terms of Hypothesis 10 the black testees perform worse than white testees in test items that involve direction and orientation with regard to direction.

The four items that were used to test this hypothesis are independent of one another and item pairs were not used as in the previous cases. The first three items in Table 5.9a are related to the application of knowledge regarding compass directions, while the fourth one, Item 56, is purely a knowledge question (cf. the items in Appendix A7). The proportions in Table 5.9a with regard to Item 56 indicate that a large percentage of the black testees did not have certain basic knowledge of compass directions: 58 % of the black testees as opposed to only 21 % of the white testees could not correctly indicate the direction in which an arrow was pointing. Items presuming this type of knowledge, for instance Item 71, are therefore a problem to black testees. The above example clearly illustrates that black testees' knowledge base with regard to compass directions is considerably less than that of white testees. The scope of Hypothesis 10 should then be interpreted as meaning that black testees can be expected to perform worse than white testees both in absolute terms and when the effect of their more limited knowledge base is taken into consideration.

The proportions in Table 5.9a reveal that the testees in all three population groups performed significantly better in Item 56 (knowledge) than in the other three items (application of the knowledge in problem solving). If the proportion for knowledge is expressed as a proportion of the total (in other words the proportion for knowledge plus the proportion for problem solving), it is also clear that the black testees not only achieved the same results as the white testees, but also performed better than the Indian testees (whites: 0,792/(0,792+0,381)= 0,67, Indians: 0,708/(0,708+0,258)= 0,73, and black: 0,418/(0,418+0,204)= 0,67). The white and black testees therefore used their knowledge equally effectively, while the Indian testees performed relatively less well than would be expected on the basis of their knowledge base. Hypothesis 10 is therefore not confirmed. According to the indices in Table 5.9b bias is present with respect to black testees in both the knowledge items

TABLE 5.9a

PROPORTION CORRECT (P), AND CONFIDENCE INTERVAL FOR P, WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR APPLIED AND THEORETICAL KNOWLEDGE OF COMPASS DIRECTIONS

| Reasoning Test: Applied and Theoretical Knowledge | Item no. | White |  | Indian |  | Black |  |
|---|---|---|---|---|---|---|---|
|  |  | p | 95 % confidence interval for p | P | 95 % confidence interval for p | p | 95 % confidence interval for p |
| Applied | 47 | 0,305 |  | 0,228 |  | 0,167 |  |
| Applied | 71 | 0,396 |  | 0,280 |  | 0,164 |  |
| Applied | 84 | 0,443 |  | 0,267 |  | 0,281 |  |
| Mean proportion-Applied: |  | 0,381 | 0,350-0,412[a] | 0,258 | 0,227-0,289[a] | 0,204 | 0,173-0,235[a] |
| -Theoretical: | 56 | 0,792 | 0,761-0,823 | 0,708 | 0,677-0,739 | 0,418 | 0,387-0,449 |

Note: [a] Differences between the two proportions significant ($p<0,05$).

TABLE 5.9b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.9a

| Description | Item no. | Bias indices |  |  |  |
|---|---|---|---|---|---|
|  |  | $\chi^2_3$ |  | ICC |  |
|  |  | Indian | Black | Indian | Black |
| Applied knowledge | 47 | 5,8 | 9,4* | 0,09 | a |
| Applied knowledge | 71 | 12,2* | 20,4* | 0,16 | 0,23+ |
| Applied knowledge | 84 | 41,5* | 24,5* | 0,47+ | a |
| Theoretical knowledge | 56 | 2,0 | 13,4* | 0,09 | 0,32+ |

Note: *$p<0,05$
+$ICC>0,2$
[a] Not possible to estimate item parameters.

and the items in which knowledge is applied. From this one can conclude that the bias is probably caused by a lack of basic knowledge about compass directions. As far as the Indian testees are concerned, bias was limited to problem-solving items (according to the $\chi_3^2$ index). Only Item 84 was identified as biased by both indices. This item consists of a relatively long verbal description and even many bright testees apparently did not understand it.

### 5.2.5 Mathematical "story" problems and calculations

#### Hypothesis 11

According to Hypothesis 11 black testees find mathematical "story" problems more difficult than calculations.

The relevant results appear in Tables 5.10a and 5.10b and the items in Appendix A8. Table 5.10a indicates that all three population groups found numerical calculations considerably easier than the corresponding story problem in which the testee has to decide for himself what type of calculation is required to solve the problem. Although these results confirmed Hypothesis 11, they were not confined to black testees.

It was however interesting to note the problems that black testees experienced with Item 64 and particularly with Item 85: only 12 % of the testees answered these two items correctly (as opposed to the 52 % of the white and 45 % of the Indian testees who answered Item 85 correctly). This indicates that a large percentage of the black testees were unable to follow a verbal numerical argument (Item 64) and did not have the numerical skills required to solve a somewhat complex (in other words more complex than Item 16) mathematical problem (Item 85).

The black testees' problems are also reflected in the bias indices in Table 5.10b: apparently only Item 16 - which requires simple numerical calculations - is unbiased.

TABLE 5.10a

ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$ ) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR MATHEMATICAL "STORY" PROBLEMS AND CALCULATIONS

| Reasoning Test: Mathematical "Story" problems and Calculations | Item no. | White | | | | Indian | | | | Black | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ |
| Story | 7 | 67,1 | 0,279 | 0,248-0,310[a] | 686,0** | 111,3 | 0,201 | 0,170-0,232[a] | 796,1** | 36,4 | 0,163 | 0,132-0,194[a] | 733,6** |
| Calculation | 16 | | 0,963 | 0,932-0,994 | | | 0,969 | 0,938-0,999 | | | 0,878 | 0,847-0,909 | |
| Story | 64 | 2,27 | 0,325 | 0,294-0,356[a] | 96,2** | 2,71 | 0,236 | 0,205-0,267[a] | 112,8** | 0,98 | 0,125 | 0,094-0,156[b] | 0,0 |
| Calculation | 85 | | 0,525 | 0,494-0,556 | | | 0,457 | 0,426-0,488 | | | 0,126 | 0,095-0,157 | |
| Mean proportion - Story: | | | 0,302 | 0,271-0,333[a] | | | 0,219 | 0,188-0,250[a] | | | 0,144 | 0,113-0,175[a] | |
| - Calculation: | | | 0,744 | 0,713-0,775 | | | 0,713 | 0,682-0,744 | | | 0,502 | 0,471-0,533 | |

Note: [a] Differences between the two proportions significant ($p<0,05$).
[b] Differences between the two proportions not significant ($p>0,05$).

** p<0,01

* p<0,05

TABLE 5.10b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.10a

| Description | Item no. | Bias indices | | | |
|---|---|---|---|---|---|
| | | $\chi^2_3$ | | ICC | |
| | | Indian | Black | Indian | Black |
| Story | 7 | 9,7* | 17,7* | 0,14 | 0,66+ |
| Calculation | 16 | 0,1 | 0,3 | 0,17 | 0,08 |
| Story | 64 | 6,4 | 14,2* | 0,07 | 0,18 |
| Calculation | 85 | 1,8 | 56,1* | 0,03 | 0,34+ |

Note: *p<0,[illegible]
+ICC>0,2

### 5.2.6 The use of statements in test items

#### Hypothesis 12

According to Hypothesis 12 black testees perform relatively worse in test items in which a statement is made than in other types of verbal items.

The Reasoning Test contained five statement type items (cf. Appendix A9). To determine whether the black testees performed worse in this type of item than in other types of verbal reasoning items, it was necessary to establish a certain standard with regard to the difficulty values of these "other" items. This standard was obtained by using the performance of the white testees as point of departure: for each of the five statements two verbal items with more or less the same difficulty value as the statement concerned were selected from the list of item difficulty values in Appendix B (Items 10 and 48 for Statement 4, 6 and 20 for Statement 21, 14 and 17 for Statement 41, 18 and 55 for Statement 60 and 59 and 68 for Statement 78). The mean proportions of the two sets of items in Table 5.11a reveal that

- no differences were found in the case of white testees (the deliberate selection of the "other" items was thus successful),
- the Indian testees found the statements somewhat more difficult than the other item types (however these differences were not significant), and
- relatively large differences were identified among the black testees: only 19 % of these testees answered the statement type items correctly as opposed to the 29 % who were successful with the other type of verbal reasoning items.

It can be concluded from the above that approximately 10 % fewer black testees are able to correctly answer an item in the form of a statement than an item with another type of content. The results therefore support Hypothesis 12 to a certain extent.

In the light of these results it is not surprising that all the statement items were biased with respect to black testees (Table 5.11b). Although most of these items were biased with respect to the Indian testees as well, the indices were generally smaller and their performance was therefore not handicapped by bias to the same extent as that of the black testees.

TABLE 5.11a

PROPORTION CORRECT (P), AND CONFIDENCE INTERVAL FOR P, WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR STATEMENT-TYPE ITEMS AND "OTHER"* VERBAL ITEMS

| Reasoning Test: Statement-type and "other" verbal items | Item no. | White | | Indian | | Black | |
|---|---|---|---|---|---|---|---|
| | | p | 95 % confidence interval for p | p | 95 % confidence interval for p | p | 95 % confidence interval for p |
| Statement | 4 | 0,888 | | 0,678 | | 0,351 | |
| Statement | 21 | 0,540 | | 0,279 | | 0,078 | |
| Statement | 41 | 0,748 | | 0,652 | | 0,227 | |
| Statement | 60 | 0,596 | | 0,447 | | 0,095 | |
| Statement | 78 | 0,488 | | 0,291 | | 0,231 | |
| Mean proportion - Statement: | | 0,652 | 0,621-0,683[b] | 0,469 | 0,438-0,500[b] | 0,196 | 0,165-0,227[a] |
| - Other*: | | 0,652 | 0,621-0,683 | 0,512 | 0,481-0,543 | 0,294 | 0,263-0,325 |

Note: [a] Differences between the two proportions significant (p<0,05).
[b] Differences between the two proportions not significant (p>0,05)
* Items 6, 10, 14, 17, 18, 20, 48, 55, 59 and 68.

TABLE 5.11b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE STATEMENT ITEMS IN TABLE 5.11a

| Description | Item no. | Bias indices | | | |
|---|---|---|---|---|---|
| | | $\chi^2_3$ | | ICC | |
| | | Indian | Black | Indian | Black |
| Statement | 4 | 20,4* | 59,9* | 0,28+ | 0,59+ |
| Statement | 21 | 56,6* | 98,6* | 0,42+ | 0,45+ |
| Statement | 41 | 14,7* | 97,2* | 0,16 | 0,34+ |
| Statement | 60 | 12,6* | 147,7* | 0,19 | 0,70+ |
| Statement | 78 | 36,3* | 17,3* | 0,27+ | 0,49+ |

Note: *p<0,05

+ICC>0,2

### 5.2.7 Synonyms and antonyms

#### Hypothesis 13

According to Hypothesis 13 black testees find synonyms easier than antonyms.

From the mean proportions in Table 5.12a it is clear that there were no significant differences between answers to synonyms and antonyms among white or black testees (the above hypothesis is therefore not confirmed). For some reason the Indian testees found synonyms more difficult than antonyms. It was noted that black testees experienced specific problems with both item types: the highest percentage of testees who answered either of the two item types correctly was only 22,5 % (Item 9 of the Reasoning Test). In contrast, 77,6 % of the white and 66 % of the Indian testees answered this item correctly. The result once again draws attention to the problems black testees experience with language.

As could be expected from the above results, all the items also displayed relatively large bias indices (Table 5.12b). Although the largest indices were found for two of the three item pairs (antonyms in the case of the black testees), it is unlikely that much value can be attached to this result owing to the small number of item pairs involved.

### 5.3 ITEM FORMAT

Although various item formats were used in the investigation, the same or similar contents were used only for certain figure series and figure analogy items. The effect of item format as such on test performance was therefore studied only with respect to series and analogies (the items concerned appear in Appendix A11).

TABLE 5.12a

ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR ANTONYMS AND SYNONYMS

| Antonyms - Synonyms | Type | Item no. | White (English-speaking) | | | | Indian | | | | Black | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ |
| Reasoning Test | Ant. | 1 | 2,36 | 0,647 | 0,616-0,678[a] | 43,8** | 1,46 | 0,628 | 0,597-0,659[a] | 21,9** | 0,40 | 0,181 | 0,150-0,212[a] | 46,8** |
| JAT 4: Synonyms | Syn. | 14 | | 0,814 | 0,783-0,845 | | | 0,714 | 0,683-0,745 | | | 0,081 | 0,050-0,112 | |
| Reasoning Test | Ant. | 9 | 0,30 | 0,776 | 0,745-0,807[a] | 86,7** | 0,18 | 0,660 | 0,629-0,691[a] | 332,8** | 0,64 | 0,225 | 0,194-0,256[a] | 16,5** |
| JAT 4: Synonyms | Syn. | 20 | | 0,512 | 0,481-0,543 | | | 0,263 | 0,232-0,294 | | | 0,162 | 0,131-0,193 | |
| Reasoning Test | Ant. | 18 | 0,77 | 0,511 | 0,480-0,542[b] | 4,6* | 0,73 | 0,329 | 0,298-0,360[b] | 14,1** | 1,25 | 0,147 | 0,116-0,178[b] | 5,8* |
| JAT 4: Synonyms | Syn. | 22 | | 0,450 | 0,419-0,481 | | | 0,268 | 0,237-0,299 | | | 0,185 | 0,154-0,216 | |
| Mean proportion - Antonyms: | | | | 0,645 | 0,614-0,676[b] | | | 0,539 | 0,508-0,570[a] | | | 0,184 | 0,153-0,215[b] | |
| - Synonyms: | | | | 0,592 | 0,561-0,623 | | | 0,415 | 0,384-0,446 | | | 0,143 | 0,112-0,174 | |

Note: [a]Differences between the two proportions significant ($p<0,05$).
[b]Differences between the two proportions not significant ($p>0,05$).
** $p<0,01$
* $p<0,05$

TABLE 5.12b

BIAS INDICES ($\chi^2_3$ AND ICC) FOR THE ITEMS IN TABLE 5.12a

| Description | Item no. | Bias indices | | | |
|---|---|---|---|---|---|
| | | $\chi^2_3$ | | ICC | |
| | | Indian | Black | Indian | Black |
| Antonyms - Synonyms | | | | | |
| Antonym | 1 | 8,4* | 68,9* | 0,20 | 0,95+ |
| Synonym | 14 | 3,8 | 229,4* | -[1)] | - |
| Antonym | 9 | 15,0* | 154,0* | 0,41+ | 1,09+ |
| Synonym | 20 | 121,4* | 84,3* | - | - |
| Antonym | 18 | 77,3* | 122,6* | 0,48+ | 0,63+ |
| Synonym | 22 | 39,0* | 40,2* | - | - |

Note: *$p<0,05$
+ICC>0,2
1)ICC index not calculated for synonyms.

Hypothesis 14

According to Hypothesis 14 figure series and figure analogies (a) differ from each other with regard to their effect on test performance and (b) affect the test performance of the various population groups differently.

From the odds ratio, the confidence intervals and the chi-square test in Table 5.13a it is clear that, regardless of their population groups, the testees found figure series easier than similar items formulated as analogies. Hypothesis 14(a) is therefore confirmed. As the difference between the mean proportions of the two item formats was more or less of the same for white, Indian and black testees (white: 0,662/(0,662+0,477)= 0,58, Indian: 0,545/(0,545+0,389)= 0,58, and black: 0,367/(0,367+0,277)= 0,57), the conclusion can be drawn that the population groups reacted similarly to specific members of the item pairs. Hypothesis 14(b) is therefore not confirmed.

Although the series format is generally one of the best item formats to use with black testees, it is also clear that the complexity of the series plays a decisive role. For example, most black testees found Item 19, which involves the rotation and increase of the number of dots, relatively easy - as can be seen from the fact that 64,7 % of them answered it correctly. In contrast, Item 58, in which one figure "grows" at the expense of the other one, was found to be very difficult and only 10,7 % of these testees could answer it correctly. Although item 58 includes no language or any other troublesome elements, its bias index with respect to black testees is still large (Table 5.13b). What is more , the white and Indian testees also found this item considerably more difficult than Item 19 (prompting the conclusion that Item 58 has a heavier "g" load). These findings raise an important question: is the deviance of this item as reflected by its bias index the result of item bias as previously defined, or should it be attributed to inability on the part of the testees?

TABLE 5.13a

ODDS RATIO, PROPORTION CORRECT (P), CONFIDENCE INTERVAL FOR P, AND CHI-SQUARE TEST ($\chi^2$) WITH REGARD TO DIFFERENCES IN THE PROPORTIONS FOR FIGURE SERIES AND FIGURE ANALOGIES

| Reasoning Test: Figure Series - Figure Analogies | Item no. | White | | | | Indian | | | | Black | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ | Odds ratio | p | 95 % confidence interval for p | $\chi^2$ |
| F. Series | 19 | 0,36 | 0,867 | 0,836-0,898[a] | 92,0** | 0,48 | 0,766 | 0,735-0,797[a] | 70,4** | 0,58 | 0,647 | 0,616-0,678[a] | 46,7** |
| F. Analogy | 29 | | 0,701 | 0,670-0,732 | | | 0,612 | 0,581-0,643 | | | 0,517 | 0,486-0,548 | |
| F. Series | 39 | 0,28 | 0,717 | 0,686-0,748[a] | 195,1** | 0,41 | 0,576 | 0,545-0,607[a] | 110,6** | 0,56 | 0,347 | 0,316-0,378[a] | 36,3** |
| F. Analogy | 49 | | 0,417 | 0,386-0,448 | | | 0,357 | 0,326-0,388 | | | 0,231 | 0,200-0,262 | |
| F. Series | 58 | 0,67 | 0,401 | 0,370-0,432[a] | 22,3** | 0,60 | 0,292 | 0,261-0,323[a] | 32,0** | 0,74 | 0,107 | 0,076-0,138[b] | 4,3* |
| F. Analogy | 66 | | 0,312 | 0,281-0,343 | | | 0,197 | 0,166-0,228 | | | 0,082 | 0,051-0,113 | |
| Mean proportion - Figure Series: | | | 0,662 | 0,631-0,693[a] | | | 0,545 | 0,514-0,576[a] | | | 0,367 | 0,336-0,398[a] | |
| - Figure Analogies: | | | 0,477 | 0,446-0,508 | | | 0,389 | 0,358-0,420 | | | 0,277 | 0,246-0,308 | |

Note: [a] Differences between the two proportions significant ($p<0,05$).
[b] Differences between the two proportions not significant ($p>0,05$).
** $p<0,01$
* $p<0,05$

TABLE 5.13b

BIAS INDICES ($\chi_3^2$ AND ICC) FOR THE ITEMS IN TABLE 5.13a

| Description | Item no. | Bias indices | | | |
|---|---|---|---|---|---|
| | | $\chi_3^2$ | | ICC | |
| | | Indian | Black | Indian | Black |
| F. Series | 19 | 4,2 | 2,6 | 0,35+ | 0,19 |
| F. Analogy | 29 | 3,0 | 3,0 | 0,10 | 0,15 |
| F. Series | 39 | 9,7* | 23,0* | 0,17 | 0,30+ |
| F. Analogy | 49 | 5,4 | 8,6* | 0,17 | 0,63+ |
| F. Series | 58 | 9,9* | 31,7* | 0,08 | 0,77+ |
| F. Analogy | 66 | 14,3* | 21,1* | 0,14 | a |

Note: *$p<0,05$

+ICC>0,2

[a] Not possible to estimate item parameters.

## 5.4 INFORMATION IN ITEM DISTRACTORS

Test compilers seldom construct item distractors in an arbitrary way. Because distractors often reflect the typical reasoning mistakes made by testees, the selection of distractors is a valuable source of information about the test behaviour of testees. According to the item responses in Appendix B black testees display certain forms of test behaviour that do not promote good test performance. In this regard it appears that black testees, to a greater extent than white or Indian testees, tend to

- select illogical answers,
- adopt an uncritical attitude,
- ignore contradictions, and
- pay only selective attention to the facts in an item.

### 5.4.1 Illogical answers

The black testees tended to select the most illogical or worst answers from the five possibilities more often than the testees from the other two population groups. In the case of Item 66 for example (the item appears in Appendix A11), 34 % of the black testees as opposed to 14 % of the Indian and only 4 % of the white testees selected Distractor D, which is the least acceptable of the five possible answers. Although Distractor A of Item 35 (Appendix A5) was a poor choice, it was nevertheless selected by 31 % of the black testees, as opposed to 11 % of the Indian and 4 % of the white testees. Other examples in this regard are Items 54 and 69.

### 5.4.2 Uncritical attitude

The responses to Item 31 (not given in the appendix), for example, seemed to indicate that a large percentage of the black testees were not critical but merely selected the first answer that came to hand. Item 31 reads as follows:
Which statement describes a THIEF the best?
Twenty-four per cent of the black and 2 % of the white testees selected Distractor A ("Someone who is in prison"), as opposed to 82 % of the white and only 27 % of the black testees who selected the "best" answer, namely "Someone who takes someone else's property".

Item 73 (a linear syllogism) is a further illustration of the uncritical attitude of black testees in particular. According to this item

> Pete and Will together have just as much money as Nic and John together. Will and Nic have the same amount of money.
> Consequently....

This item is in the form of

$$A+B=C+D$$
$$B=C$$

Therefore A = D.

It is interesting to note that 21 % of the black as opposed to 10 % of the Indian and 9 % of the white testees selected a distractor that amounted to B>D; a further 32 % of the black, 27 % of the Indian and 13 % of the white testees maintained that A+D=2(B+C).

### 5.4.3 Ignoring contradictions

It seems that many black testees ignore contradictions or are not aware that their answers do not correspond with the information provided in the item. To mention a few examples: despite the clear statement in Item 5 (Appendix A4) that "Some divers are sailors", 20 % of the black, 5 % of the Indian and 3 % of the white testees maintained that "All divers are sailors". Although Item 60 (Appendix A9) states that "All that glitters is not silver", 39 % of the black, 9 % of the Indian and 3 % of the white testees nevertheless selected a distractor that maintains "If something glitters, it is silver"!

### 5.4.4 Selective attention

Selective attention and poor attention with regard to the information provided in an item are probably two of the most important reasons why the black testees' performance was so much poorer than that of the other two population groups in items such as analogies (the first two terms of the analogy are often ignored), series (answers are based mainly on the last two numbers in the series) and statements. The responses in Appendix B reveal that the following problems are experienced by a larger percentage of black testees than Indian and white testees:

- They do not concentrate on the information contained in the stem of the item. Item 4 (Appendix A9) is a good example here. Although this item states that "No hunter is a game warden. McPhearson is a hunter", 23 % of the black testees maintained that "Some hunters are game wardens", 17 % concluded that "McPhearson is a game warden occasionally", and a further 21 % ignored all the information and maintained that "Some game wardens are hunters".

- The testees do not use all the information provided in the stem of the item. This is very clear in a case such as Item 27 (Appendix A6), which states the following: "Quarter to twelve is to half past twelve as three o' clock is to ...?" This item consists of four terms, namely A:B::C:? From the fact that 34 % of the black testees selected a distractor such as "half past three" it can be concluded that they ignored the A term and based their answers only on the B and C terms.

- The testees base their answers on only one of the two premises in a syllogism. Despite the fact that they are told in the practice examples to take both premises or statements into account, many testees nevertheless tend to base their answers on only one. The responses to Item 44 (Appendix A4) are a typical example of this. According to this item

  "No fruit is a vegetable.
  All citrus is fruit.
  Consequently..."

It seems that 18 % of the black, 14 % of the Indian and 14 % of the white testees based their answers on only the first statement and selected a distractor such as "No vegetables are fruit", while a further 31 % of the black, 9 % of the Indian and 10 % of the white testees took only the second statement into account and selected responses such as "All fruit is citrus".

- The testees do not have the knowledge of language or linguistic nuances to interpret information figuratively rather than literally. This is illustrated very well by Item 41:
"Which conclusion can be drawn from the statement below?"
Statement: " If you do not have wings you cannot fly."
A total of 60 % of the black as opposed to 21 % of the Indian and 14 % of the white testees selected distractors that indicated a literal interpretation, namely "Ostriches have wings but cannot fly", and "Only birds can fly". Similar results were obtained for Item 51: "Which statement best explains the saying? Let sleeping dogs lie." The two distractors pointing to a literal interpretation, namely "Tired dogs need lots of sleep" and "Dogs bite if they are awakened" were selected by a total of 52 % of the black, 12 % of the Indian and only 2 % of the white testees. It is probable that items containing sayings such as the one in Item 51 are unsuitable for black testees because the reasoning process by which they are solved requires a level of familiarity with the language that these testees do not possess. To white testees this would merely be a knowledge question.

## 6.0 SUMMARY AND DISCUSSION

### 6.1 FORMULATION OF ITEMS

With respect to item formulation, the reactions of white, Indian and black testees were much more similar than expected. While certain features such as the position of the correct answer among the distractors (Hypothesis 2) had no effect on test performance, other features did have an effect. In this regard it was found that all three population groups performed considerably worse in items requiring the identification of one false or untrue answer than in similar items requiring one true answer (Hypothesis 4). A negatively formulated item was also found to be considerably more difficult than a positively formulated item (Hypothesis 5), perhaps because it deviated from the testees' everyday experience and they were not able to adapt their thought processes accordingly. The fact that both these item types were biased for black testees can probably be ascribed to language factors. The results of Hypothesis 4 are supported by Scheuneman (1980) and those of Hypothesis 5 by Scheuneman as well as Dudycha and Carpenter (1973), possibly indicating a generalizable principle that should be taken into account by test compilers, particularly in the South African situation. Although the test performance of all three population groups was adversely affected by the above item formulations, the performance of black testees was even more seriously affected because they had to cope with numerous additional problems that were not experienced to the same extent by the white and Indian testees.

As concerns group differences it appeared that the white testees benefited more than the Indian or black testees from information subtly provided in the stem of the item (Hypothesis 1). White testees have a double advantage: not only are they better able to understand the item owing to their superior knowledge of the language, but - as Plake and Huntley (1984) found with respect to certain testees - they are also able to use subtle grammatical cues in the stem and distractors of the item, thus improving their performance. Another difference between the three population groups was that they did not react similarly to the item distractors, in other words they did not find the same distractors equally attractive (Hypothesis 3). The finding that black testees tend to select less likely answers to a greater extent than white or Indian testees corresponds with the typical response pattern often found among children from lower socio-economic strata (cf. Eells *et al*. 1951 ; Wolf 1978).

The above findings have certain implications for the development of a common test for the different population groups. Although the test compiler cannot raise the economic level of the underprivileged group, he can at least try to narrow the gap between the test performances of the low and high SES groups by paying special attention to the way in which items are formulated. For example, items should be formulated positively, the "untrue" type of answer should be eliminated by asking which of the answer possibilities is "true", cues should not be given and the invention of confusing, misleading or deliberately attractive item distractors should be avoided.

## 6.2 ITEM CONTENT

As far as item content is concerned, it was expected that concrete and abstract items would not only differ in the effect they had on test performance, but that the population groups would differ with respect to their reactions to these two item types (Hypothesis 6). It was also expected that, because they are raised in a more protective and conforming environment than white testees, black testees would be more field dependent and would tend to prefer functions of the right hemisphere of the brain (cf. Kaufman 1979), meaning that they would perform better in concrete than in abstract items (Hypothesis 7). The fact that none of these expectations were confirmed suggests in the case of white and Indian testees that their performance is not affected by which content type is used. In the case of the black testees the matter is more complex because the item format used to test the hypotheses with regard to concrete and abstract content, namely syllogisms, was not at all suitable for these testees. This conclusion is unconditionally confirmed by the bias indices: According to the ICC method almost all twelve of the syllogism items were biased with respect to the black testees, while only four biased items (two concrete and two abstract) were identified for the Indian testees. As a result of the problems experienced with the item format used, it is not possible to say whether the black testees performed better in concrete than in abstract items.

The reason why the syllogism is not a suitable item format for black testees is probably the same as the one identified by Haggard (1954) with regard to low SES children, namely that such items are too academic, artificial and foreign to these pupils' everyday experience. On the basis of the findings of an experiment described by Brislin (1983) it can also be concluded that if syllogisms are formulated in an African language and the black testees are given enough practice, their performance will improve considerably. For the present however syllogisms, except perhaps for those comprising concrete-meaningful material, should be avoided in a common test for white and black testees. Although all the syllogism types were generally suitable for Indian testees, it seems that the concrete-meaningless, concrete-unreal and abstract contents were somewhat less suitable than, for example, concrete-meaningful material.

Verbal and non-verbal

According to Hypothesis 8 the population groups were expected to differ with regard to the way in which their performance was influenced by verbal and non-verbal items. This hypothesis was studied with the aid of two item formats, analogies and series. The verbal items were presented as verbal analogies and letter series and the non-verbal items as figure analogies and number series. The two content types were compared within a specific format, in other words verbal analogies were compared with figure analogies and letter series with number series. The hypothesis concerned was to some extent confirmed, with regard to the verbal analogies as opposed to figure analogies, but not with regard to letter series as opposed to number series. As far as the black testees' reaction to analogies is concerned it seemed that their performance was not improved by the mere replacement of words with figures. The bias indices clearly reveal that an equal number of verbal and figure analogies were biased. Nevertheless the largest indices were found in respect of verbal analogies, supporting the findings of researchers such as Scheuneman (1980, 1981). Verbal analogies also displayed a greater incidence of bias than figure analogies in the case of the Indian testees.

As far as series are concerned the population groups not only reacted similarly, but the two types of series (letters and numbers) did not differ from one another with regard to their effect on test performance. Irrespective of specific content type series items are probably a relatively suitable format for both Indian and black testees and few biased items were found for either of the two population groups. This finding, admittedly based on a limited number of items, therefore does not support Scheuneman's (1981) observation that bias is found more often in letter series than in other types of series. A possible reason why Hypothesis 8 is supported for analogies but not for series, is that letter series items were wrongly classified as a verbal instead of a non-verbal item type. The fact that letters are used in a series does not necessarily make the item "verbal". Basically there is no real difference between a letter series and a number series as both rely on an ability to count.

The significant role of language in item bias is clear from the percentage of biased items (significant $\chi_3^2$ values, $p<0,01$) found among the verbal and the non-verbal items of the Reasoning Test as a whole. The 85 items of the Reasoning Test consist of 59 verbal and 26 non-verbal

87

items (7 of which consist of numbers, 4 of letters and 15 of figures). In the case of the black testees 45 (that is 76%) of the verbal items displayed significant bias indices, as compared to only 14 (or 54%) of the non-verbal items. This means that 24 % of all verbal items as opposed to 46 % of all non-verbal items were unbiased. Similar trends were found for the Indian testees, except that the percentages of biased non-verbal items were even smaller than in the case of blacks: 34 (that is 58%) of the verbal items and only 4 of the non-verbal items (in other words 15 %) displayed significant bias indices. In other words: 42 % of the verbal items as opposed to 85 % of the non-verbal items were unbiased.

*The above illustrates very clearly that language is certainly the greatest single cause of item bias for both Indian and black testees.*

However in the case of the black testees language was not the only cause of bias and it does not explain, for example, why more items were biased for them than for the Indian testees, namely 59 as opposed to 38 items. In view of the fact that three times as many biased non-verbal items were identified for black as opposed to Indian testees (14 as opposed to 4 items respectively), it can be concluded that factors other than language played a role in the performance of black testees.

Synonyms and antonyms

On the basis of findings reported in the literature that was studied (for instance Scheuneman 1980, 1981), it was expected that the black testees in particular would find antonyms more difficult than synonyms. The testing of Hypothesis 13 did not support these expectations - for neither the black nor the white testees. Although white and black testees found synonyms and antonyms more or less equally difficult, there was a remarkable difference between the performances of the two groups. An average of 64 % of the white testees answered the antonym items correctly as opposed to 18 % of the black testees. With regard to the synonyms the mean percentages were 59 % and 14 % respectively for the two groups. In addition to the limited number of items that were used (with more items the findings could have been reversed), a possible reason for the result obtained for the black testees is that the small proportion of testees who answered either of the two item types correctly left little room for the two percentages to differ significantly from each other. The fact that the black testees found both item types equally difficult, that only a small percentage of them could answer either of the two item types correctly and that both types were equally biased, emphasizes the problems black testees experience with vocabulary items. Under these circumstances it is not really meaningful to study a hypothesis such as the above.

At this stage it is not clear why on average 12 % more Indian testees answered antonym items than synonym items correctly (53 % for antonyms as opposed to 41 % for synonyms). Considering the limited number of items it is possible that the results are coincidental. The fact that almost all the items were biased for these testees also indicates that they have problems with vocabulary items.

## Time concepts

In Hypothesis 9 it was postulated that the black testees would perform relatively less well than the white testees in test items in which time concepts play a role. This supposition was based on an earlier (and very tentative) observation that black technikon students apparently experience problems with test items containing time concepts (Owen 1986). The underlying rationale can be found in Rohrl's (1979) view that one of the cultural correlates for functions of the right hemisphere of the brain in certain communities is that the members of such a community have little need for watches, calendars, etc. Therefore if the black testees display a tendency towards right-hemispheric functions, they may be expected to perform relatively poorly in items containing time concepts.

The results revealed that Hypothesis 9 cannot be accepted, in other words compared with the white testees the black testees did not perform worse in time than in "other" concepts. However this result draws attention to another very important matter, namely the danger of an *ex post facto* approach in identifying biased items. If only items containing time concepts had been involved and, as was done in the present investigation, they had not been contrasted with similar items that did not involve time concepts, the conclusion would necessarily have been that time concept items are problematic for black testees. From the fact that on average only 29 % of the black testees answered the time concept items correctly and that all these items displayed large ICC index values (Table 5.8b) it can rightly be concluded that time concepts were responsible for the bias. However the fact that similar "other" items had also been answered correctly by an average of only 25 % of the testees and also displayed large bias indices provides the necessary perspective.

Because bias in the case of black testees was found with respect to items with and without time concepts and the same item format had been used for the various item pairs, it should be concluded that there was no single or main cause of bias in these items. The bias must rather be ascribed to a combination of item format and item content effects. A study of the item difficulty values (p values) in Table 5.8a reveals that (for the black testees)

- contents that were expressed in series (for example Items 26 and 36) functioned better than contents given in an analogy format (for example Items 27 and 37), and
- the story type item (format and content) led to very poor performance regardless of whether numerical values were expressed in numbers or in words ( for example Items 70 and 63 ) or whether the problem involved concepts of time or money (for example Items 46 and 55)

As far as the story type item is concerned the underlying problem (bias factor) was definitely language: if the testee did not understand the context and interrelationships of the words, it was further immaterial whether the content was expressed in numbers or in words. The fact that the black testees performed equally well in analogy items expressing the numbers in words (Item 27) and in figures (Item 37) supports the conclusion that the story element in the item causes the most problems. Although some of the testees were apparently not very familiar with the new way of expressing time (as could be seen from the difference between the proportions of testees who answered Items 26 and 36 correctly), the impact of this problem cannot be compared with that of language difficulties.

As regards the Indian testees the time concept items functioned satisfactorily in general.

### Mathematical story problems

In investigating the above hypothesis (Hypothesis 9) it was concluded that language - and specifically the story element - was responsible for the problems experienced by black testees with arithmetical or mathematical story problems. The role of language ability, compared with numerical ability, in this type of item was studied further by means of Hypothesis 11. This hypothesis suggested that the black testees find dealing with mathematical story problems more difficult than performing the calculations concerned.

Although the results supported the hypothesis, they also applied to the white and Indian testees. This finding indicates that inherently it is more difficult to follow the story of a calculation problem than to perform the numerical calculations involved - probably because the solution of the former item type requires more than one ability (Reckase 1985). A further analysis of the items indicates that a large percentage of the black testees firstly were unable to follow a verbal-numerical argument (on the basis of the findings with regard to Hypothesis 9 it can be assumed that language ability played a considerable role in this regard), and secondly did not have the numerical skills and knowledge required to solve a more complex mathematical problem (in other words a problem that demands more than the absolutely basic skills of Item 16, for example). The bias indices indicate that only the simplest numerical items are unbiased for black testees. No biased items were found for the Indian testees. The bias results for the black testees imply certain problems for the test compiler: in the case of story problems the bias could perhaps be decreased by manipulating the language, but numerical calculations, where the bias is due to a lack of basic skills that the majority of testees should have, leave the test compiler almost no room for manoeuvring.

### Direction and orientation

The testing of Hypothesis 10 revealed that, considering their respective knowledge bases, the black testees did not perform worse than their white counterparts in items concerning direction and direction orientation. However the crux of the matter is that the knowledge base of the black testees was considerably more limited than that of the white and Indian testees, as was revealed by the fact that a large percentage (58 %) of these testees were unable to indicate correctly the direction in which an arrow pointed. Items based on knowledge of compass directions are therefore inevitably biased with respect to black testees. Although one can only speculate at this stage about black testees' lack of elementary knowledge with regard to compass directions represented on paper, it is reasonable to assume that their school and general reading experience in this regard is more limited than that of white and Indian testees. The fact that the black testees had problems orienting themselves with regard to a specific direction represented on paper could indicate field dependence, but there is insufficient information to successfully pursue this thought.

The most important point emphasized by Hypothesis 10 is that, if the test compiler wants to narrow the gap in test performance between white and black testees, he should not include items that involve compass directions in tests that primarily measure reasoning ability.

### Statements

The testing of Hypothesis 12 revealed that the black testees performed worse in verbal reasoning items in which a statement was made than in other types of verbal items. All the statement items also indicated quite large bias indices. Most of the statement items were also biased for the Indian testees, but the indices were generally smaller than for the black testees.

The fact that on average 10 % fewer black testees managed to answer a statement item correctly than another verbal item type can naturally not be attributed to language problems (after all the language is similar for the two types of item content). A more probable explanation is that the same factors that lead to poor performance in syllogisms are involved. In other words, statement type items are too academic, artificial and foreign to the everyday world of black testees.

Owing to a lack of practice and experience black testees are not used to reasoning within the strict confines of a statement and limiting their thoughts to the facts given in the statement. Seen from another angle, it is not impossible that persons who tend to think holistically - which is often associated with right-hemispheric brain functions (cf. Kaufman 1979) - will experience problems with statement type items (and syllogisms) that demand analytical thinking.

On the other hand, Hale's (1982) assertion that African blacks regard a direct statement as rough and unimaginative cannot be completely ignored. A black testee, for example, will definitely not be motivated to go to any trouble over an item that he regards as an unintelligent verbal construction. However further information in this regard cannot be obtained from a group test situation and all the possible reasons given for poor performance with regard to statements must therefore be regarded as only tentative.

## 6.3 ITEM FORMAT

Only two item formats, namely figure series and figure analogies were formally compared. The same item content was used so as to determine the effect of the item format on test performance. The results with regard to Hypothesis 14 revealed that all three population groups found figure series easier than figure analogies. The extent to which the same item content is made more difficult by the format in which it is presented is clear from the mean proportions (x100=%) in Table 5.13a. For example, an average of 66,2 % of the white testees answered the figure series correctly, while the corresponding percentage for the figure analogies was 47,7 % (a difference of almost 19 %). In the case of the Indian testees the corresponding percentages were 54,5 % and 38,9 % (a difference of 15,6 %), and for the black testees 36,7 % and 27,7 % (a difference of 9 %). This finding corresponds with the view of Sternberg and Gardner (1983:90) that

> *...one would expect that for problems in these formats with terms of equal difficulty (i.e. drawn from the same universe of stimuli), analogies would be slightly more difficult than series completion...*

According to these writers series are easier than analogies because fewer information-processing parameters are required to solve series than to solve analogies.

Although series is undoubtedly one of the best item formats for use with black testees, its suitability depends on the complexity or degree of difficulty of the item concerned: the more difficult the item (as measured by the performance of white testees), the less suitable it is for black testees. This matter is illustrated very clearly by the bias indices in Table 5.13b, where two of the three figure series items display reasonably large values (the more difficult the item, the greater the values - this applies to both the $\chi^2_3$ and ICC methods). This finding highlights a basic question raised throughout the investigation, namely whether deviating items are the result of bias - according to the definition at the end of Paragraph 3.2 - or whether they should be ascribed to the inability of the testees. Although no definite answer can be given to this question, the way in which the testees worked to solve the problems - as concluded from their choice of item distractors - nevertheless provides some clarity.

As far as the Indian testees are concerned neither series nor analogies gave many problems.

## 6.4 INFORMATION IN ITEM DISTRACTORS

From the responses to item distractors it seems that the black testees, to a greater extent than the white or Indian testees, tended to

select illogical answers;
adopt an uncritical attitude;
ignore contradictions, and
pay selective attention to the facts in an item:

- They did not keep to the information provided in the stem of the item.
- They did not use all the information provided in the stem of the item.
- They based their answers on only one of the two statements of a syllogism.

It was also clear that they did not have the necessary linguistic knowledge and finer understanding of language nuances required to interpret certain data figuratively rather than literally ( sayings items are therefore not suitable for black testees). A lack of arithmetical knowledge and skills also played a large role in calculation problems.

The above forms of test behaviour, which are reasonably typical of test-unsophisticated persons from a low SES, contributed greatly to the identification of deviant and therefore biased items. In a previous report (Owen 1989) about the same testees the above aspects of test behaviour were identified as causes of apparent item bias - although the statistical indices indicate that an item is biased, the real causes of the bias are not located in the item as such, but rather in the test behaviour of the testees.

One of the most important causes of true item bias identified in the investigation is the role of language (as a bias factor) in verbal constructions such as word classification, linear syllogisms, verbal analogies and mathematical story problems. Owing to language problems the black testees were unable to reveal their true potential with respect to the above item types. Any test consisting mainly of such items will necessarily be biased for these testees.

## 7.0 RECOMMENDATIONS

To make the compilation of a common test battery for white and black testees in South Africa a reality, the considerable gap that presently exists between the test performance of the two groups will have to be narrowed. In the case of reasoning tests (verbal and non-verbal) the matters mentioned below with respect to the formulation of items, item content and item format could contribute towards better test performances by black testees. Owing to the complex nature of the problems these "solutions" are in no way complete and other factors such as the mastering of thinking strategies, cognitive enrichment, practice and training will also have to receive attention.

Unfortunately a multiple aptitude test battery cannot consist only of series, no matter how much one would like to use this most suitable item format. Language is indispensable to the measurement of the full spectrum of intellectual abilities required in a technological society. Until the proficiency of black testees in one of the official languages improves, the test compiler will always come up against bias caused by language deficiencies. However a start can be made towards the establishment of common tests by applying the following rules:

### Item formulation

The following should be avoided as far as possible:

- Negatively formulated item stems and answer possibilities
- Items requiring one false or untrue answer
- Item distractors that are overly attractive and confusing to less sophisticated testees
- Subtle cues in the stem and/or distractors of the item.

### Item content

The following should be avoided:

- Difficult words in items such as word classification and analogies
- Complicated sentence constructions in mathematical story problems
- The use in reasoning tests of items involving direction and orientation with respect to direction (Geography tests is a different matter)
- Synonyms and/or antonymous words in tests that do not mainly measure language matters
- Test items containing statements
- The involvement of basic arithmetic skills in items that are not primarily concerned with measuring numerical skills
- Test items containing sayings.

If the test compiler has a choice between verbal analogies and figure analogies, he should choose the latter.

### Item format

The following two formats should be avoided as far as possible:

- Syllogisms (not at all suitable)
- Linear syllogisms (suitability depending to a certain extent on the complexity of the story element).

Series of any content type (letters, numbers or figures) is possibly the best format. Analogies are less suitable but cannot always be replaced by series without affecting the construct validity of the test.

In conclusion it should be pointed out that, due to the limited number of items used in testing the various hypotheses, the findings of this report should be regarded as guidelines, and as exploratory rather than definitive. It is also clear that further information on the thought processes of testees is required for a cohesive explanation of their test behaviour. In order to make progress in this regard attention will have to be given to the approach and method of cognitive psychologists in their study of information-processing procedures.

## 8.0 REFERENCES

Abrams, N.E. 1979. A comparison of performance of black and white students on tests of racially orientated information. Unpublished PH.D. Dissertation, Columbia University. (Univ. Microfilms International, Ann Arbor, Michigan, 1982)

Anastasi, A. 1970. On the formation of psychological traits. American Psychologist, 25, 899-910.

Anastasi, A. 1976. Psychological testing. New York: MacMillan, Fourth Edition.

Anastasi, A. 1985. Some emerging trends in psychological measurement: A fifty-year perspective. Applied Psychological Measurement, 9(2), 121-138.

Angoff, W.H. 1982. Use of difficulty and discrimination indices for detecting item bias. In R.A. Berk(Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Bond, L. 1981. Bias in mental tests. New Directions for Testing and Measurement, 11, 55-77.

Breland, H.M., Stocking, M., Pinchak, B.& Abrams, N. 1974. The cross-cultural stability of mental test items: An investigation of response patterns for ten socio-cultural groups. Report PR-74-2. Princeton, N.J.: Educational Testing Service.

Brislin, R.W. 1983. Cross-cultural research in psychology. Annual Review of Psychology, 34, 363-400.

Burrill, L.E. 1981. A comparative investigation into the identification of ethnic bias in items assessing current educational status. Unpublished PH.D. Dissertation, Fordham University, New York. (Univ. Microfilms International, Ann Arbor, Michigan, 1984)

Burrill, L.E. & Wilson, R. 1980. Fairness and the matter of bias. Test Service Notebook 36. New York: The Psychological Corporation.

Buss, A.R. 1977. On the relationship between the psychological environment and the development of differences in abilities. Intelligence, 1, 192-207.

Carlton, S.T. & Marco, G.L. 1982. Methods used by test publishers to "debias" standardized tests: Educational Testing Service. In R.A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Cleary, T.A. & Hilton, T.L. 1968. An investigation of item bias. Educational and Psychological Measurement, 28, 61-75.

Coffman, W.E. 1982. Methods used by test publishers to "debias" standardized tests: Riverside Publishing Company / Houghton Mifflin. In R.A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Cole, M.& Scribner, S. 1974. Culture and thought: A psychological introduction. New York: John Wiley.

Coop, R.H. & Sigel, I.E. 1971. Cognitive style: Implications for learning and instruction. Psychology in the Schools, 8(2), 152-161.

Donlon, T.F., Hicks, M.M. & Wallmark, M.M. 1980. Sex differences in item responses on the Graduate Record Examination. Applied Psychological Measurement, 4(1), 9-20.

Dubb, A.A. 1974. The impact of the city. In W.D. Hammond-Tooke(Ed.), The Bantu-speaking peoples of Southern Africa. London: Routledge & Kegan Paul.

Dudycha, A.L. & Carpenter, J.B. 1973. Effects of item format on item discrimination and difficulty. J. of Applied Psychology, 58, 116-121.

Ebel, R.L. 1979. Intelligence: a skeptical view. J.of Research and Development in Education, 12(2), 14-21.

Eells, K.W., Davis, A., Havighurst, R.J., Herrick, V.E. & Tyler, R.W. 1951. Intelligence and cultural differences. Chicago: University of Chicago Press.

Federico, P.-A. 1980. Adaptive instruction: Trends and issues. In R.E. Snow, P.-A. Federico & W.E. Montague(Eds.), Aptitude, learning, and instruction. Vol.1: Cognitive process analyses of aptitude. Hillsdale, N.J.: Lawrence Erlbaum.

Federico, P.-A. 1983. Changes in the cognitive components of achievement as students proceed through computer-managed instruction. J.of Computer-Based Instruction, 9(4), 156-168.

Ferguson, G.A. 1954. On learning and human ability. Canadian J. of Psychology, 8(2), 95-112.

Fishbein, R.L. 1975. An investigation of the fairness of items of a test battery. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Washington, D.C.,March 31-April 2. (ERIC Document Reproduction Service No. ED 111 837)

Forsyth, R.A. & Spratt, K.F. 1980. Measuring problem solving ability in mathematics with multiple-choice items: The effect of item format on selected item and test characteristics. J. of Educational Measurement, 17(1), 31-43.

Frederiksen, N. 1984. The real test bias: Influences of testing on teaching and learning. American Psychologist, 39(3), 193-202.

Ghuman, P.A.S. 1980. A Study of the concept of equivalence and divergent thinking among four sub-cultural groups of Punjabi children. International Review of Applied Psychology, 29(1/2), 89-103.

Globerson, T., Weinstein, E. & Sharabany, R. 1985. Teasing out cognitive development from cognitive style: A training study. Developmental Psychology, 21(4), 682-691.

Green, D.R. 1982. Methods used by test publishers to "debias" standardized tests: CTB / McGraw-Hill. In R.A. Berk(Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Guilford, J.P. 1980. Cognitive styles: What are they? Educational and Psychological Measurement, 40(3), 715-735.

Guthrie, G. 1963. Structure of abilities in a non-western culture. J. of Educational Psychology, 54(2), 94-103.

Haggard, E.A. 1954. Social-status and intelligence: An experimental study of certain cultural determinants of measured intelligence. Genetic Psychology Monograph, 49, 141-186.

Hale, J.E. 1982. Black children: Their roots, culture and learning styles. Provo, Utah: Brigham Young University Press.

Hambleton, R.K., Swaminathan, H., Cook, L.L., Eignor, D.R. & Gifford, J.A. 1978. Developments in latent trait theory: Models, technical issues, and applications. Review of Educational Research, 48(4), 467-510.

Handrick, F.A. & Loyd, B.H. 1982. Methods used by test publishers to "debias" standardized tests: The American College Testing Program.In R.A. Berk(Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Haney, W. 1984. Testing reasoning and reasoning about testing. Review of Educational Research, 54(4), 597-654.

Havighurst, R.J. & Breese, F.H. 1947. Relation between ability and social status in a Midwestern community. III Primary mental abilities. J. of Educational Psychology, 38, 241-247.

Horn, J.L. 1976. Human abilities: A review of research and theory in the early 1970s. Annual Review of Psychology, 27, 437-485.

Hughes, H.H. & Trimble, W.E. 1965. The use of complex alternatives in multiple choice items. Educational and Psychological Measurement, 25, 117-126.

Humphreys, L.G. 1986. An analysis and evaluation of test and item bias in the prediction context. J. of Applied Psychology, 71(2), 327-333.

Humphreys, L.G. & Taber, T. 1973. Ability factors as a function of advantaged and disadvantaged groups. *J. of Educational Measurement*, 10(2), 107-115.

Hunt, E. 1980. The foundations of verbal comprehension. In R.E. Snow, P.-A. Federico & W.E. Montague(Eds.), *Aptitude, learning and instruction. Vol.1: Cognitive process analyses of aptitude*. Hillsdale, N.J.: Lawrence Erlbaum.

Hunt, E. 1982. Towards new ways of assessing intelligence. *Intelligence*, 6, 231-240.

Ironson, G.H. 1982. Use of chi-square and latent trait approaches for detecting item bias. In R.A. Berk(Ed.), *Handbook of methods for detecting test bias*. Baltimore: The Johns Hopkins University Press.

Ironson, G.H. & Subkoviak, M.J. 1979. A comparison of several methods of assessing item bias. *J. of Educational Measurement*, 16(4), 209-225.

Irvine, S.H. 1970. Affect and construct -- a cross-cultural check on theories of intelligence. *J.of Social Psychology*, 80, 23-30.

Jensen, A.R. 1969. How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review*, 39, 1-123.

Jensen, A.R. 1974. How biased are culture-loaded tests? *Genetic Psychology Monographs*, 90, 185-244.

Jensen, A.R. 1980. *Bias in mental testing*. London: Methuen.

Kaufman, A.S. 1979. Cerebral specialization and intelligence testing. *J. of Research and Development in Education*, 12(2), 96-107.

Kulick, E.& Dorans, N.J. 1984. *The standardization approach to assessing unexpected differential item performance*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 23-27. (ERIC Document Reproduction Service No. ED 247 260)

Lenke, J.M. 1982. Methods used by test publishers to "debias" standardized tests: The Psychological Corporation. In R.A. Berk (Ed.), *Handbook of methods for detecting test bias*. Baltimore: The Johns Hopkins University Press.

Lesser, G.S., Fifer, G. & Clark, D.H. 1965. Mental abilities of children from different social-class and cultural groups. *Monographs of the Society for Research in Child Development*, 30(4), 1-115.

Linn, R.L., Levine, M.V., Hastings, C.N. & Wardrop, J.L. 1980. *An investigation of item bias in a test of reading comprehension*. Technical Report No. 163. Center for the study of reading. Champaign, Illinois: University of Illinois at Urbana- Champaign. (ERIC Document Reproduction Service No. ED 184091)

Linn, R.L., Levine, M.V., Hastings, C.N. & Wardrop,J.L. 1981. Item bias in a test of reading comprehension. *Applied Psychological Measurement*, 5(2), 159-173.

Linn, R.L. & Harnisch, D.L. 1981. Interactions between item content and group membership on achievement test items. *J. of Educational Measurement*, 18(2), 109-118.

Lord, F.M. 1977. A study of item bias, using item characteristic curve theory. In Y.H.Poortinga(Ed.), *Basic problems in cross-cultural psychology*. Amsterdam: Swets & Zeitlinger.

Lorge, I. 1952/53. Intelligence and cultural differences as seen by the test maker. *Teachers College Record*, 54, 190-193.

MacKenzie, A.J. 1981. Level I and level II abilities in primary school children. *British J. of Educational Psychology*, 51, 312-320.

Marascuilo, L.A. & Slaughter, R.E. 1981. Statistical procedures for identifying possible sources of item bias based on $X^2$ statistics. *J. of Educational Measurement*, 18(4), 229-248.

Marmorale, A.M. & Brown, F. 1979. Psychological testing of the non-white, nonmiddle class child. *International J. of Group Tensions*, 1-4, 195-200.

McGee, M.G. 1979. Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, 86(5), 889-918.

Medley, D.M. & Quirk, T.J. 1972. *Race and subject-matter influences on performance on general educational items of the National Teacher Examinations*. Report RB.-72-43. Princeton, N.J.: Educational Testing Service.

Medley, D.M. & Quirk, T.J. 1974. The application of a factorial design to the study of cultural bias in general culture items on the National Teacher Examinations. *J. of Educational Measurement*, 11, 235-245.

Mellenbergh, G.J. 1983. Conditional item bias methods. In S.H.Irvine & J.W. Berry(Eds.), *Human assessment and cultural factors*. New York: Plenum Press.

Merryfield, M.M. 1985. The challenge of cross-cultural evaluation: Some views from the field. *New Directions for Program Evaluation*, 25, 3-17.

Metropolitan Achievement Tests 1978. *Steps used to eliminate ethnic "bias"*. Special Report No. 16. New York: The Psychological Corporation.

Nagy, P. & Griffiths, A.K. 1982. Limitations of recent research relating Piaget's theory to adolescent thought. *Review of Educational Research*, 52(4), 513-556.

Novick, M.R. 1980. Policy issues of fairness in testing. In L.J.T. van der Kamp, W.F.Langerak & D.N.M. de Gruijter(Eds.), *Psychometrics for educational debates*. Chichester: John Wiley.

Oakland, T. 1977. *Psychological and educational assessment of minority children*. New York: Brunner/Mazel.

Oakland, T. & Feigenbaum, D. 1979. Multiple sources of test bias on the WISC-R and Bender-Gestalt Test. *J of Consulting and Clinical Psychology*, 47(5), 919-927.

Olmedo, E.L. 1981. Testing linguistic minorities. *American Psychologist*, 36(10), 1078-1085.

Osterlind, S.J. 1983. *Test item bias*. Beverly Hills: Sage Publications.

Owen, K. 1986. *Toets- en itemsydigheid: Toepassing van die Senior Aanlegtoetse, Meganiese Insigtoets en Skolastiese Bekwaamheidsbattery op blanke, swart, kleurling- en Indiërtechnikonstudente* (Test and item bias: Application of the Senior Aptitude Tests, Mechanical Insight Test and Scholastic Proficiency Battery to white, black, coloured and Indian technikon students). Report P-66. Pretoria: Human Sciences Research Council.

Owen, K. 1989. *Test and item bias: The suitability of the Junior Aptitude Tests as a common test battery for white, Indian and black pupils in Standard 7.* Report P-96. Pretoria: Human Sciences Research Council.

Pellegrino, J.W. & Glaser, R. 1980. Components of inductive reasoning. In R.E. Snow, P.-A. Federico & W.E. Montague(Eds.), *Aptitude, learning, and instruction. Vol.1: Cognitive process analyses of aptitude*. Hillsdale, N.J.: Lawrence Erlbaum.

Petersen, N.S. 1980. Bias in the selection rule-bias in the test. In L.J.T. van der Kamp, W.F. Langerak & D.N.M. de Gruijter(Eds.), *Psychometrics for educational debates*. Chichester: John Wiley.

Plake, B.S. 1980. A comparison of a statistical and subjective procedure to ascertain item validity: One step in test validation process. *Educational and Psychological Measurement*, 40(2), 397-404.

Plake, B.S. & Hoover, H.D. 1979/80. An analytical method of identifying biased test items. *The J. of Experimental Education*, 48(2), 153-154.

Plake, B.S. & Huntley, R.M. 1984. Can relevant grammatical cues result in invalid test items? *Educational and Psychological Measurement*, 44(3), 687-696.

Radford, J & Burton, A. 1972. Changing intelligence. In K.Richardson, D. Spears & M. Richards(Eds.), *Race, culture and intelligence*. Middlesex, England: Penguin Books.

Raju, N.S. 1982. Methods used by test publishers to "debias" standardized tests: Science Research Associates. In R.A. Berk(Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Reckase, M.D. 1985. The difficulty of test items that measure more than one ability. Applied Psychological Measurement, 9(4), 401-412.

Ree, M.J. 1979. Estimating item characteristic curves. Applied Psychological Measurement, 3(3), 371-385.

Reschly, D.J. 1984. Aptitude tests. In G. Goldstein & M. Hersen(Eds.), Handbook of psychological assessment. New York: Pergamon Press.

Rohrl, V.J. 1979. Culture, cognition and intellect: Towards a cross-cultural view of intelligence. The J. of Psychological Anthropology, 2(3),337-364.

Rudner, L.M. 1978. Using standard tests with the hearing impaired: The problem of item bias. The Volta Review, 80(1), 31-40.

Rudner, L.M., Getson, P.R. & Knight, D.L. 1980. Biased item detection techniques. J. of Educational Statistics, 5(2), 213-233.

Scarr, S. 1981. Testing for children: Assessment and the many determinants of intellectual competence. American Psychologist, 36(10), 1159-1166.

Scheuneman, J.D. 1978. Further considerations in the assessment of bias in test items. Paper presented at the Meeting of the American Psychological Association, Toronto.

Scheuneman, J.D. 1979. A new method of assessing bias in test items. J. of Educational Measurement, 16(3), 143-152.

Scheuneman, J.D. 1980 Latent-trait theory and item bias. In L.J.T. van der Kamp, W.F. Langerak & D.N.M. de Gruijter(Eds.), Psychometrics for educational debates. Chichester: John Wiley.

Scheuneman, J.D. 1981. A new look at bias in aptitude tests. New Directions for Testing and Measurement, 12, 3-35.

Scheuneman, J.D. 1982. Item bias and test scores. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, March.

Scheuneman, J.D. 1983. Bias in test items: Causes and effects. Paper presented at the Annual Meeting of the American Psychological Association, Anaheim, California, August.

Scheuneman, J.D. 1985. Exploration of causes of bias in test items. Research Report 85-42. Princeton, N.J.: Educational Testing Service.

Schmeiser, C.B. 1982. Use of experimental design in statistical item bias studies. In R.A.Berk(Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Schultz, L.J. & Fortune, J.C. 1981. The three I's: Sources of test bias. Education, 102(2), 117-123.

Shepard, L.A. 1981. Identifying bias in test items. New Directions for Testing and Measurement, 11, 79-104.

Shepard, L.A., Camilli, G. & Averill, M. 1981. Comparison of procedures for detecting test-item bias with both internal and external ability criteria. J. of Educational Statistics, 6(4), 317-375.

Shepard, L.A., Camilli, G. & Williams, D.M. 1983. Accounting for statistical artifacts in item bias research. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, April.

Snow, R.E. 1980. Aptitude process. In R.E. Snow, P.-A.Federico & W.E. Montague(Eds.), Aptitude, learning and instruction. Vol.I: Cognitive process analyses of aptitude. Hillsdale, N.J.: Lawrence Erlbaum.

Sternberg, R.J. & Gardner, M.K. 1983. Unities in inductive reasoning. J. of Experimental Psychology: General, 112(1), 80-116.

Stricker, L.J. 1982. Identifying test items that perform differentially in population subgroups: A partial correlation index. Applied Psychological Measurement, 6(3), 261-273.

Subkoviak, M.J., Mack, J.S., Ironson, G.H. & Craig, R.D. 1984. Empirical comparison of selected item bias detection procedures with bias manipulation. J. of Educational Measurement, 21(1), 49-58.

Sundberg,N.D. & Gonzales, L.R. 1981. Cross-cultural and cross-ethnic assessment: Overview and issues. In P.McReynolds(Ed.), Advances in psychological assessment. Vol.5. San Francisco: Jossey-Bass.

Tatsuoka, K.K.& Linn, R.L. 1983. Indices for determining unusual patterns: Links between two general approaches and potential applications. Applied Psychological Measurement, 7, 81-96.

Templer, A.J. 1972. The relationship between field dependence-independence and concept attainment. Psychologia Africana, 14,121-129.

Thorndike, R.L. 1982. Applied psychometrics. Boston: Houghton Mifflin.

Thorndike, R.L. & Hagen, E. 1969. Measurement and evaluation in psychology and education. New York: John Wiley.

Tittle, C.K. 1982. Use of judgmental methods in item bias studies. In R.A.Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Triandis, H.C. & Brislin, R.W. 1984. Cross-cultural psychology. American Psychologist, 39(9), 1006-1016.

Valencia, R.R. & Rankin, R.J. 1985. Evidence of content bias on the McCarthy Scales with Mexican American children: Implications for test translation and nonbiased assessment. J. of Educational Psychology, 77(2), 197-207.

Van der Vliet, V. 1974. Growing up in traditional society. In W.D. Hammond-Tooke(Ed.), The Bantu-speaking peoples of Southern Africa. London: Routledge & Kegan Paul.

Vernon,P.E. 1968. What is potential ability? Bulletin of the British Psychological Society, 21, 211-219.

Vernon, P.E. 1969. Intelligence and cultural environment. London: Methuen.

Walberg, H.J. & Haertel, G. 1984. Community influences on learning. Evaluation in Education: An International Review Series, 8, 1-82.

Wigdor, A.K. & Garner, W.R. 1982. Ability testing: Uses, consequences and controversies. Part I and II. Documentation Section. Washington D.C.: National Academy Press.

Wilcox, R.R. 1984/85. A note on measuring item bias. The J. of Experimental Education, 53(2), 114-116.

Williamson, M.L. & Hopkins, K.D. 1967. The use of "none of these" versus homogeneous alternatives on multiple-choice tests: Experimental reliability and validity comparisons. J. of Educational Measurement, 4(2), 53-58.

Witelson, S.F. 1977. Developmental dyslexia: Two right hemispheres and none left. Science, 195, 309-311.

Witkin, H.A., Goodenough, D.R. & Karp, S.A. 1967. Stability of cognitive style from childhood to young adulthood. J. of Personality and Social Psychology, 7, 291-300.

Witkin, H.A., Moore, C.A., Goodenough, D.R. & Cox, P.W. 1977. Field-dependent and field-independent cognitive styles and their educational implications. Review of Educational Research, 47, 1-64.

Wolf, B. 1978. Bias in testing. Paper presented at the 19th International Congress of Applied Psychology.

Wood, R.L., Wingersky,M.S. & Lord, F.M. 1976. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. Research Memorandum RM-76-6(modified 1/78). Princeton, N.J.:Educational Testing Service.

## APPENDIX A

## REASONING TEST

**TABLE A1: CUES IN TEST ITEMS**

| Items contain cue | Items without cue |
|---|---|
| 13. A statement which can have more than one meaning has got a(n) ... meaning.?<br>A. distorted.<br>B. erroneous.<br>C. far-fetched<br>D. improper.<br>E. double. | 22. Evidence, which can be interpreted wrongly, has a(n) ... meaning.<br>A. distorted<br>B. erroneous<br>C. far-fetched<br>D. improper<br>E. double |
| 32. Someone who has been influenced in making a decision based on preconceived opinions, is said to be ...<br>A. influential.<br>B. hypocritical.<br>C. prejudiced.<br>D. accusatory.<br>E. impartial. | 42. A man who accuses somebody else before he has all the facts at his disposal is said to be ...?<br>A. influential.<br>B. hypocritical.<br>C. prejudiced.<br>D. occusatory.<br>E. impartial. |

TABLE A2: TRUE-FALSE ANSWERS

| One true answer | One false answer |
|---|---|
| 3. Carlos is younger than Fred but older than Nico. Frank is just as old as Carlos. Which one of the following statements is true?<br>A. Fred is younger than Nico.<br>B. Frank is older than Nico.<br>C. Carlos is younger than Frank.<br>D. Nico is older than Carlos.<br>E. Frank is older than Fred. | 11. Mark is shorter than Will but taller than John. Steve is just as tall as Mark. Which one of the following statements is false?<br>A. Will is taller than John.<br>B. Steve is shorter than John.<br>C. Mark is just as tall as Steve.<br>D. John is shorter than Mark.<br>E. Will is taller than Steve. |
| 30. Four people stand in a row. Jack stands in front of Peter and Peter stands in front on Sam. Martin stands in front of Jack. Which one of the following statements is true?<br>A. Jack stands right in the front.<br>B. Peter stands second from the front.<br>C. Martin stands third from the back.<br>D. Sam stands right at the back.<br>E. Peter stands in front of Martin. | 40. Four people get into a bus. Mich gets into the bus before Gert and Gert before Peter. Gene gets into the bus before Mich.<br>Which one of the following statements is false?<br>A. Peter got into the bus after Gert.<br>B. Gert got into the bus after Mich.<br>C. Peter got into the bus last.<br>D. Mich got into the bus first.<br>E. Gene got into the bus before Mich. |

TABLE A3: POSITIVE-NEGATIVE TEST ITEMS

| Positive | Negative |
|---|---|
| 3. Carlos is younger than Fred but older than Nico. Frank is just as old as Carlos. Which one of the following statements is true?<br>A. Fred is younger than Nico.<br>B. Frank is older than Nico.<br>C. Carlos is younger than Frank.<br>D. Nico is older than Carlos.<br>E. Frank is older than Fred. | 20. Nic has less money than Jane but more then Louise. Linn has just as much money as Nic. Which one of the following statements is true?<br>A. Jane does not have more money than Louise.<br>B. Linn does not have less money than Louise.<br>C. Linn does not have less money than Jane.<br>D. Louise does not have less money than Nic.<br>E. Jane does not have more money than Nic. |
| 5. All divers are swimmers.<br>Some divers are sailors.<br>Consequently:<br>A. Some sailors are swimmers.<br>B. Some swimmers are not divers.<br>C. All sailors are swimmers.<br>D. All divers are sailors.<br>E. Only swimmers are sailors. | 44. No fruit is a vegetable.<br>All citrus is fruit.<br>Consequently:<br>A. No citrus is a vegetable.<br>B. No vegetables are fruit.<br>C. Some vegetables are not fruit.<br>D. All fruit is citrus.<br>E. It is not possible to reach a valid conclusion. |

**TABLE A3:** (Continued)

| Positive | Negative |
|---|---|
| 30. Four people stand in a row. Jack stands in front of Peter and Peter stands in front on Sam. Martin stands in front of Jack. Which one of the following statements is true?<br>A. Jack stands right in the front.<br>B. Peter stands second from the front.<br>C. Martin stands third from the back.<br>D. Sam stands right at the back.<br>E. Peter stands in front of Martin. | 50. Four people stand in a queue in a butchery. Betty is served before Anne and Anne before Susan. Jean is served before Betty. Which one of the following statements is true?<br>A. Susan was not served last.<br>B. Anne was not served after Betty.<br>C. Jean was not served first.<br>D. Betty was not served before Jean.<br>E. Susan was not served after Anne. |
| 59. I do not have the most pocket-money but I have more than my brother Pete. My sister Anne has less pocket-money than Pete. My brother Ben also has some pocket-money. Which one of the following statements is true if the amounts of our pocket-money differ?<br>A. Anne has more pocket-money than I.<br>B. I have more pocket-money than Ben.<br>C. Ben has the most pocket-money.<br>D. Pete has the least pocket-money.<br>E. Ben has less pocket-money than Anne. | 67. I am not the eldest but I am older than my brother Ian. My sister Mary is not older than Ian. I also have a brother called Sam. Which one of the following statements is not true if our date of births differ?<br>A. Mary is not older than I.<br>B. I am not older than Sam.<br>C. Sam is not the eldest.<br>D. Ian is not the youngest.<br>E. I am not younger than Mary. |

TABLE A4: CONCRETE-ABSTRACT TEST ITEMS

| Concrete | Abstract |
|---|---|
| 5. All divers are swimmers.<br>Some divers are sailors.<br>Consequently:<br>A. Some sailors are swimmers.<br>B. Some swimmers are not divers.<br>C. All sailors are swimmers.<br>D. All divers are sailors.<br>E. Only swimmers are sailors. | 14. All D are S.<br>Some D are M.<br>Consequently:<br>A. Some M are S.<br>B. Some S are D.<br>C. All M are S.<br>D. All D are M.<br>E. Only S are M. |
| 24. All animals with long hair have four legs. All animals with four legs have spots.<br>Consequently:<br>A. All animals with spots have four legs.<br>B. All animals with four legs have long hair.<br>C. All animals with spots have long hair.<br>D. All animals have spots.<br>E. All animals with long hair have spots. | 34. All Wurps with Zug have Jabsy.<br>All Wurps with Jabsy have Mesbet.<br>Consequently:<br>A. All Wurps with Mesbet have Jabsy.<br>B. All Wurps with Jabsy have Zug.<br>C. All Wurps with Mesbet have Zug.<br>D. All Wurps have Mesbet.<br>E. All Wurps with Zug have Mesbet. |

TABLE A4: (Continued)

| Concrete | Abstract |
|---|---|
| 44. No fruit is a vegetable.<br>All citrus is fruit.<br>Consequently:<br>A. No citrus is a vegetable.<br>B. No vegetables are fruit.<br>C. Some vegetables are not fruit.<br>D. All fruit is citrus.<br>E. It is not possible to reach a valid conclusion. | 53. No A are B.<br>All C are A.<br>Consequently:<br>A. No C are B.<br>B. No B are A.<br>C. Some B are not A.<br>D. All A are C.<br>E. It is not possible to reach a valid conclusion. |
| 61. No pumpkins are vegetables.<br>All potatoes are pumpkins.<br>Consequently:<br>A. No vegetables are pumpkins.<br>B. Some vegetables are not pumpkins.<br>C. No pumpkins are potatoes.<br>D. No potatoes are vegetables.<br>E. It is not possible to reach a valid conclusion. | 68. No Zerps are Oktras.<br>All Iskis are Zerps.<br>Consequently:<br>A. No Oktras are Zerps.<br>B. Some Oktras are not Zerps.<br>C. No Zerps are Iskis.<br>D. No Iskis are Oktras.<br>E. It is not possible to reach a valid conclusion. |

TABLE A4: (Continued)

| Concrete | Abstract |
|---|---|
| 74. All monkeys have wings.<br>No zebras have wings.<br>Consequently:<br>A. Only monkeys have wings.<br>B. No zebras are monkeys.<br>C. All monkeys are zebras.<br>D. No monkeys have wings.<br>E. Some zebras do not have wings. | 79. All A have W.<br>No Z have W.<br>Consequently:<br>A. Only A have W.<br>B. No Z are A.<br>C. All A are Z.<br>D. No A have W.<br>E. Some Z do not have W. |
| 81. Some carrots are hares.<br>Some hares can fly.<br>Consequently:<br>A. All carrots can fly.<br>B. Some carrots are not hares.<br>C. Some hares cannot fly.<br>D. All hares have long ears.<br>E. It is not possible to reach a valid conclusion. | 83. Some A are F.<br>Some F are K.<br>Consequently:<br>A. All A are K.<br>B. Some A are not F.<br>C. Some F are not K.<br>D. All F are letters in the alphabet.<br>E. It is not possible to reach a valid conclusion. |

**TABLE A5: VERBAL - NON-VERBAL TEST ITEMS**

| (i) Letter series | Number series |
|---|---|
| 28. a b c d e f g h i j k l m n o p q r s t u v w x y z<br>What is the next letter in the sequence c; e; g; i; ...?<br>A. j<br>B. o<br>C. l<br>D. k<br>E. another letter | JAT2: Item no. 2* |
| 10. a b c d e f g h i j k l m n o p q r s t u v w x y z<br>The groups of letters in the box go together in a certain way. Find the group of letters that goes where you see the question mark (?) in the box.<br>c; d; e    d; e; f    e; f; g<br>l; m; n    m; n; o    ?<br>A. h; i; j<br>B. n; o; p<br>C. j; k; l<br>D. m; n; p<br>E. o; p; q | 2. The groups of numbers in the box go together in a certain way. Find the group of numbers that goes where you see the question mark (?) in the box.<br>3; 4; 5    4; 5; 6    5; 6; 7<br>12; 13; 14    13; 14; 15    ?<br>A. 8; 9; 10<br>B. 14; 15; 16<br>C. 10; 11; 12<br>D. 13; 14; 16<br>E. 15; 16; 17 |

*See JAT test booklet

TABLE A5: (Continued)

| (i) Letter series | Number series |
|---|---|
| 48. a b c d e f g h i j k l m n o p q r s t u v w x y z<br>What is the next letter in the sequence a; f; b; g; c; ...?<br>A. d<br>B. e<br>C. j<br>D. h<br>E. another letter | 38. What is the next number in the sequence 1; 6; 2; 7; 3; ...?<br>A. 4<br>B. 5<br>C. 10<br>D. 8<br>E. another number |
| 65. a b c d e f g h i j k l m n o p q r s t u v w x y z<br>The groups of letters in the box go together in a certain way. Find the group of letters that goes where you see the question mark (?) in the box.<br>a; d; f    d; g; i    f; i; k<br>i; l; n    l; o; q    ?<br>A. n; q; s<br>B. m; p; r<br>C. o; r; t<br>D. n; p; s<br>E. o; q; s | 57. The groups of numbers in the box go together in a certain way. Find the group of numbers that goes where you see the question mark (?) in the box.<br>1; 4; 6    4; 7; 9    6; 9; 11<br>9; 12; 14    12; 15; 17    ?<br>A. 14; 17; 19<br>B. 13; 16; 18<br>C. 15; 18; 20<br>D. 14; 16; 19<br>E. 15; 17; 19 |

TABLE A5: (Continued)

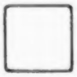
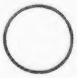
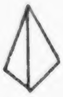
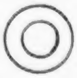
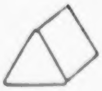
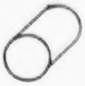
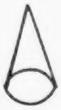
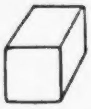
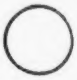
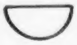
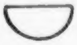
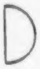
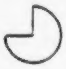
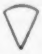
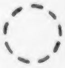
| (ii)Verbal Analogies | Figure Analogies |
|---|---|
| 6. SQUARE is to CIRCLE as PYRAMID is to ...?<br>A. NILE<br>B. CIRCUMFERENCE<br>C. BALL<br>D. CONE<br>E. CUBE | 15.  is to  as  is to ...?<br>A. <br>B. <br>C. <br>D. <br>E.  |
| 25. FULL is to HALF as HALF is to ...?<br>A. THIRD<br>B. QUARTER<br>C. EIGHTH<br>D. THREE-QUARTERS<br>E. EMPTY | 35.  is to  as  is to ...?<br>A <br>B <br>C <br>D <br>E  |

TABLE A5: (Continued)

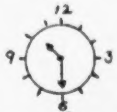
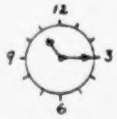
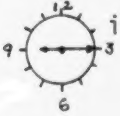
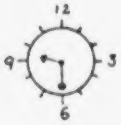
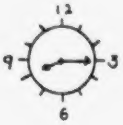
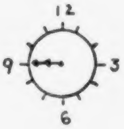
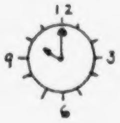
| (ii)Verbal Analogies | Figure Analogies |
|---|---|
| 27. *Quarter to twelve* is to *Half past twelve* as *Three o'clock* is to ...?<br>A. Half past three<br>B. Quarter past two<br>C. Half past one<br>D. Quarter to four<br>E. None of these | 37.  is to  as  is to ...?<br>A. <br>B. <br>C. <br>D. <br>E. None of these |

TABLE A5: (Continued)

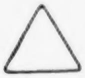
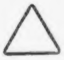
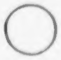
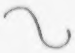
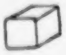
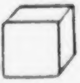
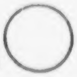
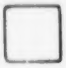
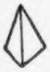
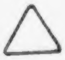
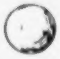
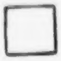
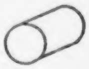
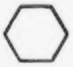
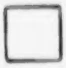
| (ii)Verbal Analogies | Figure Analogies |
|---|---|
| 45. SQUARE is to LINE as LINE is to ... ?<br>A. POINT<br>B. TRIANGLE<br>C. SURFACE<br>D. ROPE<br>E. BLOCK | 54.  is to —— as —— is to ...?<br>A <br>B <br>C <br>D <br>E  |
| 62. What is related to a CUBE in the same way a CIRCLE is related to a SQUARE?<br>A. CONE<br>B. TRIANGLE<br>C. BALL<br>D. SQUARE<br>E. PIPE | 69. .?. is to  as  is to <br>A <br>B <br>C <br>D <br>E  |
| 75. A PENTAGON (a figure with five angles) is to THREE as a TRIANGLE is to ...?<br>A. TWO<br>B. FOUR<br>C. THREE<br>D. FIVE<br>E. ONE | 82.  is to 4 as  is to ...?<br>A. 1<br>B. 3<br>C. 5<br>D. 4<br>E. 2 |

TABLE A6: TIME CONCEPTS

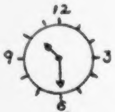
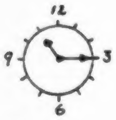
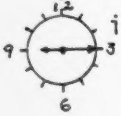
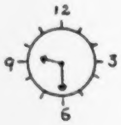
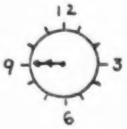
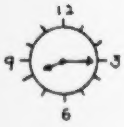
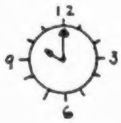
| i) Verbal | Number |
|---|---|
| 70. A train departs on Wednesday morning at ten minutes to eleven and arrives on Thursday at twenty minutes past four in the morning at its destination. How long did the journey take?<br>A. 16 hours 30 minutes<br>B. 18 hours 30 minutes<br>C. 16 hours 50 minutes<br>D. 17 hours 30 minutes<br>E. 17 hours 10 minutes | 63. A bus departs on Monday at 10 h 40 and arrives at its destination on Tuesday at 02 h 30. How long did the journey take?<br>A. 14 hours 50 minutes<br>B. 16 hours 30 minutes<br>C. 14 hours 30 minutes<br>D. 15 hours 50 minutes<br>E. 15 hours 10 minutes |
| (ii) Verbal | Figure |
| 27. *Quarter to twelve* is to *Half past twelve* as *Three o'clock* is to ...?<br>A. Half past three<br>B. Quarter past two<br>C. Half past one<br>D. Quarter to four<br>E. None of these | 37.  is to  as  is to ...?<br>A. <br>C. <br>B. <br>D. <br>E. None of these |

TABLE A6:(Continued)

| (iii)Time concept | Other concept |
|---|---|
| 8. Give one word for the following: Day, Year, Century, Week, Hour.<br>A. Date<br>B. Age<br>C. Period<br>D. Distance<br>E. Seasons | 17. Give one word for the following: Nic, Johannesburg, John, Durban, Japan.<br>A. Letters<br>B. Places<br>C. Names<br>D. Words<br>E. Cities |
| 26. What time comes next in the sequence<br>10 h 09 ; 10 h 22 ; 10 h 35 ; 10 h 48 ;...?<br>A. 10 h 51<br>B. 11 h 01<br>C. 11 h 11<br>D. 11 h 02<br>E. 11 h 00 | 36. How many litres comes next? 10,63 litres;10,76 litres;10,89 litres; ...?<br>A. 10,99 litres<br>B. 11,02 litres<br>C. 11,12 litres<br>D. 11,01 litres<br>E. 11,00 litres |

**TABLE A6:(Continued)**

| (iii)Time concept | Other concept |
|---|---|
| 46. A watch loses 4 seconds in 12 hours. If I had put it right on a Wednesday at 07 h 00, how slow would it be on the next Saturday at 07 h 00?<br>A. 18 seconds<br>B. 30 seconds<br>C. 12 seconds<br>D. 36 seconds<br>E. 24 seconds | 55. A man loses R4 on every 12 bags of onions that he sells. If he sold 3 consignments of 24 bags each, what was his total loss?<br>A. R18<br>B. R30<br>C. R12<br>D. R36<br>E. R24 |
| 76. The *Day Before Yesterday* is to *Tomorrow* as *Tuesday* is to ...?<br>A. Wednesday<br>B. Thursday<br>C. Friday<br>D. Saturday<br>E. Monday | 80. *Two School Standards Back* is to the *Next School Standard* as *Standard* 5 is to ...?<br>A. Standard 6<br>B. Standard 7<br>C. Standard 8<br>D. Standard 9<br>E. Standard 4 |

## TABLE A7: COMPASS DIRECTIONS

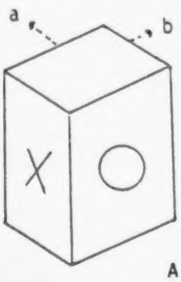
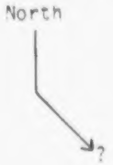
| Applied knowledge | Theoretical knowledge |
|---|---|
| 47. Mac lives 5 km south-west of Tom. Salty lives 7 km east of Mac. In which direction must Salty walk to visit Tom?<br>A. North-west<br>B. North<br>C. South-east<br>D. West<br>E. South<br>71. A man approaches a crossroad from the west. He is on his way to Maun. The road straight ahead of him ends at a river. The road that turns left goes to Tierpoort. In which direction must the man turn in order to reach Maun?<br>A. North<br>B. South<br>C. East<br>D. West<br>E. Impossible to tell<br>84. As shown on the sketch a box has a circle and a cross on it. You are standing upright in the box and are facing in the direction of a. The box is now turned 90 degrees and it faces in the direction of b. You did not move and are still facing a. On which side of you is the circle now?<br><br> <br>A. Right-hand side<br>B. Left-hand side<br>C. In front of you<br>D. Behind you<br>E. Impossible to tell | 56. In which direction does the arrow point?<br><br> <br>A. South-east<br>B. North-east<br>C. East<br>D. South<br>E. South-west |

TABLE A8: MATHEMATICAL "STORY" PROBLEMS-CALCULATIONS

| Story | Calculations |
|---|---|
| 7. I put my watch, which loses 3 minutes an hour, right at 06 h 00. When I arrived at the station at 11 h 00, according to *my* watch, the 11 h 10 train had already left. If the train had departed exactly according to the timetable, by how many minutes had I missed the train?<br>A. 10 minutes<br>B. 15 minutes<br>C. 5 minutes<br>D. 3 minutes<br>E. 2 minutes | 16. What is the answer to the following calculations? (11-6) x 3 - 10 = ?<br>A. 10<br>B. 15<br>C. 5<br>D. 3<br>E. 2 |
| 64. A man buys 30 kilograms of potatoes for R0,50 a kilogram. Six kilograms of potatoes went bad and he sold the remainder for R1,50 a kilogram. What was his profit on the whole transaction?<br>Which one of the following calculations must be made in order to solve the problem?<br>A. 1,5 x (30 - 6) - (30 x 0,5)<br>B. (30 x 1,5) - (30 x 0,5) - 6<br>C. 0,5 x (30 - 6) - (30 x 1,5)<br>D. (24 ÷ 0,5) + (30 x 1,5)<br>E. None of these. | 85. What is the answer to the following calculations:<br>(30-6) x 1,5 - (0,5 x 30) = ?<br>A. 21<br>B. 24<br>C. 27<br>D. 18<br>E. None of these. |

TABLE A9: STATEMENT-TYPE AND "OTHER" VERBAL ITEMS

Statement-type items

4. Which *one* true conclusion can be drawn from the statement below if *only the information in the statement* is taken into account?

Statement: No hunter is a game-warden. McPhearson is a hunter.

A. McPhearson makes a lot of money selling ivory.

B. Some hunters are game-wardens.

C. McPhearson is a game-warden occasionally.

D. Some game-wardens are hunters.

E. McPhearson is not a game-warden.

21. Which *one* true conclusion can be drawn from the statement below if *only the information in the statement* is taken into account?

Statement: Some plants also blossom in autumn.

A. Some plants blossom only once a year.

B. Some plants also blossom in winter.

C. Certain plants blossom only in autumn.

D. Some plants blossom more than once a year.

E. Roses blossom in autumn.

TABLE A9: (Continued)

Statement-type items

41. Which conclusion can be drawn from the statement below?

Statement: If you do not have wings you cannot fly.

A. You cannot attempt something if you do not have the means.

B. Ostriches have wings but cannot fly.

C. Only birds can fly.

D. If you are not strong you must be clever.

E. One cannot depend upon the help of another.

60. Which conclusion can be drawn from the statement below?

Statement: All that glitters is not silver.

A. Silver is dull sometimes.

B. Appearances are often deceptive.

C. Silver is expensive jewellery.

D. If something glitters, it is silver.

E. One cannot always believe what you hear.

**TABLE A9: (Continued)**

Statement-type items

78. Which *one* true conclusion can be drawn from the statement below if *only the information in the statement* is taken into account?

Statement: The world's population is increasing at a tremendous rate but food production remains constant.

A. The world is steadily becoming more crowded.

B. Few people will have enough to eat in the future.

C. People will become skinnier in the future.

D. Less food will be available for animals in the future.

E. The population becomes denser where there is a good supply of food.

**TABLE A9: (Continued)**

### Other verbal items

6. SQUARE is to CIRCLE as PYRAMID is to ...?

A. NILE

B. CIRCUMFERENCE

C. BALL

D. CONE

E. CUBE

10. a b c d e f g h i j k l m n o p q r s t u v w x y z

The groups of letters in the box go together in a certain way. Find the group of letters that goes where you see the question mark (?) in the box.

| | | |
|---|---|---|
| c; d; e | d; e; f | e; f; g |
| l; m; n | m; n; o | ? |

A. h; i; j

B. n; o; p

C. j; k; l

D. m; n; p

E. o; p; q

**TABLE A9: (Continued)**

Other verbal items

14. All D are S.
Some D are M.

Consequently:

A. Some M are S.

B. Some S are D.

C. All M are S.

D. All D are M.

E. Only S are M.

17. Give one word for the following:
Nic, Johannesburg, John, Durban,
Japan.

A. Letters

B. Places

C. Names

D. Words

E. Cities

18. What is the opposite of
*squander?*

A. promote

B. protect

C. repair

D. spend

E. save

TABLE A9: (Continued)

Other verbal items

20. Nic has less money than Jane but more then Louise. Linn has just as much money as Nic. Which one of the following statements is true?

    A. Jane does not have more money than Louise.

    B. Linn does not have less money than Louise.

    C. Linn does not have less money than Jane.

    D. Louise does not have less money than Nic.

    E. Jane does not have more money than Nic.

48. a b c d e f g h i j k l m n o p q r s t u v w x y z

    What is the next letter in the sequence a; f; b; g; c; ...?

    A. d

    B. e

    C. j

    D. h

    E. another letter

TABLE A9: (Continued)

Other verbal items

55. A man loses R4 on every 12 bags of onions that he sells. If he sold 3 consignments of 24 bags each, what was his total loss?

A. R18

B. R30

C. R12

D. R36

E. R24

59. I do not have the most pocket-money but I have more than my brother Pete. My sister Anne has less pocket-money than Pete. My brother Ben also has some pocket-money. Which one of the following statements is true if the amounts of our pocket-money differ?

A. Anne has more pocket-money than I.

B. I have more pocket-money than Ben.

C. Ben has the most pocket-money.

D. Pete has the least pocket-money.

E. Ben has less pocket-money than Anne.

TABLE A9: (Continued)

Other verbal items

68. No Zerps are Oktras.
All Iskis are Zerps.

Consequently:

A. No Oktras are Zerps.

B. Some Oktras are not Zerps.

C. No Zerps are Iskis.

D. No Iskis are Oktras.

E. It is not possible to reach a valid conclusion.

TABLE A10: ANTONYMS-SYNONYMS

| Antonyms | Synonyms (JAT 4) |
|---|---|
| 1. What is the opposite of *frequently?*<br>A. seldom<br>B. sometimes<br>C. repeatedly<br>D. often<br>E. never | Item 14* |
| 9. What is the opposite of *understand?*<br>A. forget<br>B. misapprehend<br>C. miscalculate<br>D. mistake<br>E. ignorant | Item 20 |
| 18. What is the opposite of *squander?*<br>A. promote<br>B. protect<br>C. repair<br>D. spend<br>E. save | Item 22 |

*See JAT test booklet

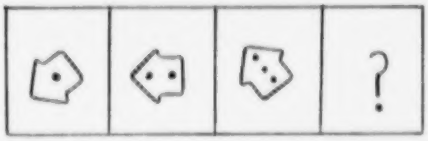
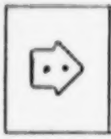
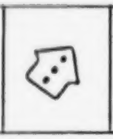
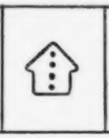
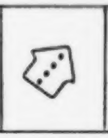
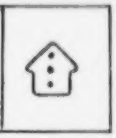
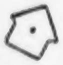
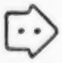
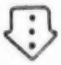
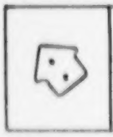
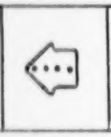
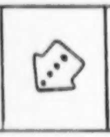
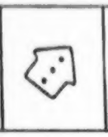
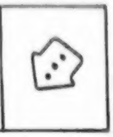
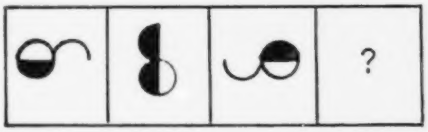
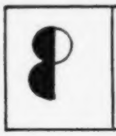
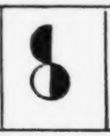
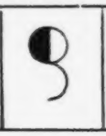
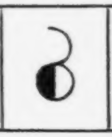
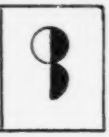
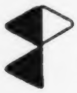
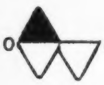
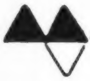
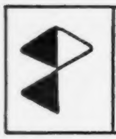
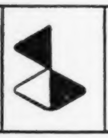
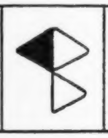
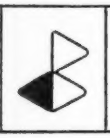
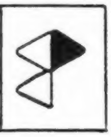
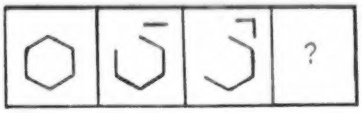
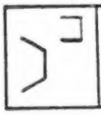
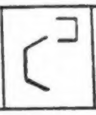
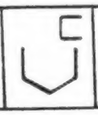
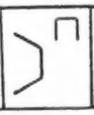
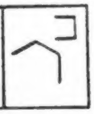
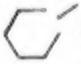
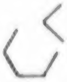
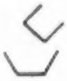
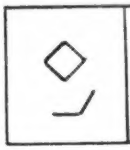
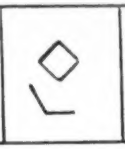
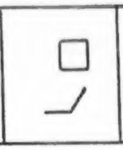
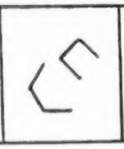
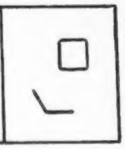
| TABLE A11 : ITEM FORMAT | |
|---|---|
| Figure Series | Figure Analogies |
| 19. Find the figure that goes where you see the question mark (?).<br><br>A <br>B <br>C <br>D <br>E  | 29.  is to  as  is to ...?<br>A <br>B <br>C <br>D <br>E  |
| 39. Find the figure that goes where you see the question mark (?).<br><br>A <br>B <br>C <br>D <br>E  | 49.  is to  as  is to ...?<br>A <br>B <br>C <br>D <br>E  |

TABLE A11: (Continued)

| Figure Series | Figure Analogies |
|---|---|
| 58. Find the figure that goes where you see the question mark (?).<br><br> <br>A <br>B <br>C <br>D <br>E  | 66.  is to  as  is to ...?<br>A <br>B <br>C <br>D <br>E  |

## APPENDIX B

## PERCENTAGE OF TESTEES WHO SELECTED A PARTICULAR OPTION AND PROPORTION CORRECT(P) FOR THE ITEMS OF THE REASONING TEST

**TABLE B:**
**REASONING TEST: PERCENTAGE OF TESTEES WHO SELECTED A PARTICULAR OPTION AND PROPORTION CORRECT (P)**

| | White | | | | | | Indian | | | | | | Black | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | % testees selecting each option | | | | | p | % testees selecting each option | | | | | p | % testees selecting each option | | | | | p |
| | 1 | 2 | 3 | 4 | 5 | | 1 | 2 | 3 | 4 | 5 | | 1 | 2 | 3 | 4 | 5 | |
| 1 | 56,4* | 6,5 | 11,6 | 14,1 | 11,3 | 0,56 | 62,8* | 5,5 | 5,4 | 19,2 | 7,2 | 0,63 | 18,1* | 17,4 | 51,1 | 7,1 | 6,3 | 0,18 |
| 2 | 1,2 | 83,6* | 0,6 | 0,8 | 13,8 | 0,84 | 3,3 | 80,0* | 1,4 | 1,3 | 13,9 | 0,80 | 11,7 | 51,6* | 5,1 | 8,0 | 23,5 | 0,52 |
| 3 | 7,7 | 68,9* | 7,7 | 8,4 | 7,2 | 0,69 | 8,5 | 53,1* | 12,0 | 13,3 | 13,1 | 0,53 | 12,1 | 30,4* | 33,4 | 11,2 | 12,9 | 0,30 |
| 4 | 2,0 | 2,7 | 2,6 | 4,0 | 88,8* | 0,89 | 2,8 | 7,9 | 9,3 | 12,2 | 67,8* | 0,68 | 4,0 | 22,6 | 16,7 | 21,4 | 35,1* | 0,35 |
| 5 | 64,8* | 18,2 | 10,8 | 3,3 | 2,8 | 0,65 | 47,3* | 22,0 | 19,0 | 4,9 | 6,9 | 0,47 | 33,9* | 20,2 | 17,3 | 19,9 | 8,6 | 0,34 |
| 6 | 14,0 | 7,6 | 5,6 | 52,6* | 20,2 | 0,53 | 22,1 | 21,1 | 4,1 | 36,4* | 16,4 | 0,36 | 9,7 | 34,4 | 13,1 | 18,4* | 24,4 | 0,18 |
| 7 | 20,5 | 44,8 | 27,9* | 5,6 | 1,3 | 0,28 | 24,6 | 40,8 | 20,1* | 12,7 | 1,8 | 0,20 | 33,9 | 23,3 | 16,3* | 21,8 | 4,8 | 0,16 |
| 8 | 21,2 | 10,2 | 63,5* | 0,8 | 4,3 | 0,64 | 13,3 | 11,9 | 69,4* | 0,6 | 4,9 | 0,69 | 22,1 | 9,3 | 44,5* | 3,9 | 20,2 | 0,45 |
| 9 | 3,8 | 81,8* | 6,4 | 3,2 | 4,7 | 0,82 | 12,0 | 66,0* | 6,0 | 4,0 | 11,9 | 0,66 | 42,8 | 22,5* | 5,7 | 8,3 | 20,7 | 0,23 |
| 10 | 1,3 | 89,5* | 0,7 | 1,3 | 7,2 | 0,90 | 2,4 | 84,1* | 0,9 | 2,5 | 10,2 | 0,84 | 14,6 | 55,1* | 4,5 | 7,3 | 18,6 | 0,55 |
| 11 | 15,3 | 45,8* | 17,8 | 10,0 | 11,1 | 0,46 | 14,9 | 39,4* | 17,7 | 12,8 | 15,3 | 0,39 | 18,8 | 23,2* | 28,4 | 15,9 | 13,7 | 0,23 |
| 12 | 14,4 | 14,0 | 15,3 | 14,0 | 42,4* | 0,42 | 18,5 | 11,6 | 18,8 | 20,4 | 30,6* | 0,31 | 23,6 | 13,9 | 26,1 | 26,6 | 9,7* | 0,10 |
| 13 | 10,3 | 13,7 | 6,0 | 4,9 | 65,1* | 0,65 | 12,9 | 34,4 | 10,1 | 14,7 | 27,9* | 0,28 | 9,3 | 28,6 | 12,2 | 21,2 | 28,7* | 0,29 |
| 14 | 73,2* | 10,8 | 8,5 | 4,1 | 3,3 | 0,73 | 57,1* | 17,4 | 13,2 | 7,8 | 4,4 | 0,57 | 28,2* | 28,6 | 16,7 | 17,2 | 9,4 | 0,28 |
| 15 | 5,6 | 24,6 | 8,1 | 48,1* | 13,5 | 0,48 | 5,5 | 39,1 | 6,4 | 33,6* | 15,4 | 0,34 | 6,3 | 38,4 | 12,2 | 23,3* | 19,8 | 0,23 |
| 16 | 0,4 | 0,6 | 96,3* | 0,5 | 2,3 | 0,96 | 0,2 | 1,0 | 96,9* | 0,7 | 1,2 | 0,97 | 1,5 | 6,1 | 87,8* | 2,0 | 2,6 | 0,88 |
| 17 | 1,4 | 1,0 | 75,0* | 20,5 | 2,1 | 0,75 | 3,4 | 6,7 | 64,4* | 21,7 | 3,8 | 0,64 | 7,3 | 16,3 | 32,8* | 23,2 | 20,5 | 0,33 |
| 18 | 7,5 | 8,6 | 5,2 | 16,8 | 61,9* | 0,62 | 12,7 | 24,2 | 10,0 | 20,2 | 32,9* | 0,33 | 23,8 | 20,6 | 16,3 | 24,6 | 14,7* | 0,15 |
| 19 | 0,7 | 0,6 | 86,7* | 8,9 | 3,1 | 0,87 | 1,0 | 1,7 | 76,6* | 15,9 | 4,7 | 0,77 | 2,7 | 4,0 | 64,7* | 17,5 | 11,1 | 0,65 |
| 20 | 10,9 | 53,4* | 11,0 | 14,7 | 10,1 | 0,53 | 13,4 | 43,6* | 11,8 | 17,4 | 13,9 | 0,44 | 15,5 | 25,8* | 20,7 | 18,0 | 20,1 | 0,26 |
| 21 | 5,3 | 6,2 | 30,7 | 54,0* | 3,8 | 0,54 | 7,7 | 10,1 | 47,0 | 27,9* | 7,4 | 0,28 | 17,5 | 14,2 | 48,9 | 7,8* | 11,6 | 0,08 |
| 22 | 42,7 | 22,9 | 6,6 | 16,3 | 11,4* | 0,11 | 27,4 | 27,8 | 11,7 | 26,6 | 6,5* | 0,07 | 17,2 | 38,1 | 11,8 | 22,1 | 10,8* | 0,11 |
| 23 | 46,0* | 3,7 | 3,4 | 3,9 | 43,0 | 0,46 | 51,7* | 8,9 | 14,7 | 16,9 | 7,8 | 0,52 | 29,1* | 20,8 | 18,0 | 23,9 | 8,2 | 0,29 |
| 24 | 24,6 | 19,3 | 21,0 | 3,3 | 31,8* | 0,32 | 28,4 | 23,1 | 18,9 | 5,2 | 24,4* | 0,24 | 29,7 | 30,9 | 14,8 | 10,2 | 14,5* | 0,15 |
| 25 | 1,6 | 44,6* | 0,6 | 1,7 | 51,5 | 0,45 | 0,9 | 69,2* | 0,8 | 5,1 | 24,0 | 0,69 | 6,3 | 50,1* | 4,0 | 5,0 | 34,7 | 0,50 |
| 26 | 8,5 | 76,4* | 2,6 | 7,7 | 4,8 | 0,76 | 13,9 | 66,9* | 4,2 | 7,1 | 7,9 | 0,67 | 36,2 | 36,5* | 6,9 | 8,3 | 12,2 | 0,37 |
| 27 | 9,3 | 1,8 | 1,1 | 72,8* | 15,0 | 0,73 | 23,0 | 2,8 | 1,0 | 59,8* | 13,4 | 0,60 | 33,8 | 4,6 | 4,8 | 31,1* | 25,7 | 0,31 |
| 28 | 3,1 | 0,2 | 1,3 | 94,6* | 0,8 | 0,95 | 4,2 | 0,5 | 2,9 | 91,4* | 1,0 | 0,91 | 11,1 | 1,5 | 5,4 | 80,4* | 1,6 | 0,80 |
| 29 | 0,3 | 25,0 | 70,1* | 2,0 | 2,6 | 0,70 | 0,9 | 30,4 | 61,2* | 4,3 | 3,1 | 0,61 | 6,0 | 26,0 | 51,7* | 8,9 | 7,4 | 0,52 |
| 30 | 9,5 | 6,0 | 4,2 | 76,8* | 3,5 | 0,77 | 17,4 | 8,8 | 11,5 | 57,3* | 5,0 | 0,57 | 30,5 | 17,8 | 20,5 | 19,7* | 11,5 | 0,20 |
| 31 | 2,4 | 1,0 | 14,0 | 81,8* | 0,8 | 0,82 | 6,0 | 2,0 | 41,2 | 45,2* | 5,6 | 0,45 | 23,7 | 14,0 | 28,2 | 27,1* | 7,1 | 0,27 |
| 32 | 52,3 | 13,9 | 21,3* | 2,4 | 10,2 | 0,21 | 50,0 | 20,9 | 15,5* | 5,3 | 8,3 | 0,16 | 47,1 | 17,5 | 17,1* | 8,8 | 9,5 | 0,17 |
| 33 | 69,9* | 15,3 | 2,0 | 0,9 | 11,9 | 0,70 | 37,4* | 26,4 | 7,2 | 2,8 | 26,2 | 0,37 | 14,9* | 43,1 | 16,8 | 12,8 | 12,4 | 0,15 |
| 34 | 29,1 | 12,8 | 17,7 | 9,4 | 31,0* | 0,31 | 32,2 | 18,2 | 19,5 | 10,5 | 19,6* | 0,20 | 29,9 | 22,0 | 16,4 | 9,8 | 21,9* | 0,22 |
| 35 | 3,7 | 71,1* | 2,3 | 2,4 | 20,5 | 0,71 | 10,9 | 69,7* | 5,4 | 2,6 | 11,4 | 0,70 | 30,7 | 39,7* | 10,9 | 5,2 | 13,5 | 0,40 |
| 36 | 5,5 | 81,5* | 4,3 | 5,5 | 3,2 | 0,82 | 10,1 | 75,4* | 5,6 | 6,6 | 2,4 | 0,75 | 29,6 | 47,2* | 8,8 | 8,4 | 6,0 | 0,47 |
| 37 | 12,0 | 2,4 | 6,3 | 70,4* | 9,0 | 0,70 | 17,2 | 5,9 | 8,9 | 58,5* | 9,5 | 0,59 | 25,9 | 11,2 | 20,6 | 30,7* | 11,6 | 0,31 |
| 38 | 1,2 | 2,9 | 1,4 | 92,0* | 2,5 | 0,92 | 3,2 | 7,2 | 2,3 | 82,7* | 4,7 | 0,83 | 14,8 | 17,2 | 7,6 | 52,2* | 8,2 | 0,52 |
| 39 | 12,3 | 10,1 | 3,2 | 2,7 | 71,7* | 0,72 | 12,3 | 16,7 | 6,7 | 6,7 | 57,6* | 0,58 | 11,2 | 21,5 | 19,0 | 13,5 | 34,7* | 0,35 |
| 40 | 12,4 | 8,5 | 12,9 | 51,3* | 14,9 | 0,51 | 12,9 | 10,2 | 15,0 | 44,9* | 17,0 | 0,45 | 19,3 | 20,0 | 20,8 | 20,5* | 19,5 | 0,21 |
| 41 | 74,8* | 3,9 | 10,1 | 5,4 | 5,8 | 0,75 | 65,2* | 10,6 | 10,3 | 2,6 | 11,3 | 0,65 | 22,7* | 25,6 | 34,0 | 7,5 | 10,1 | 0,23 |
| 42 | 1,0 | 14,9 | 33,9* | 39,2 | 10,9 | 0,34 | 6,3 | 21,8 | 29,2* | 28,1 | 14,5 | 0,29 | 15,2 | 18,1 | 17,9* | 34,4 | 14,4 | 0,18 |
| 43 | 4,0 | 0,6 | 0,6 | 6,3 | 88,6* | 0,89 | 20,9 | 2,3 | 0,8 | 10,7 | 65,4* | 0,65 | 45,4 | 13,7 | 11,8 | 12,5 | 16,6* | 0,17 |
| 44 | 57,2* | 13,5 | 5,2 | 9,5 | 14,6 | 0,57 | 46,0* | 13,7 | 12,5 | 9,4 | 18,4 | 0,46 | 19,8* | 17,8 | 23,0 | 31,0 | 8,4 | 0,20 |
| 45 | 56,7* | 9,1 | 11,7 | 6,8 | 15,6 | 0,57 | 54,9* | 10,7 | 10,6 | 10,5 | 13,4 | 0,55 | 40,1* | 23,5 | 8,8 | 12,9 | 14,8 | 0,40 |

**TABLE B: (Continued)**

| Item | White | | | | | | Indian | | | | | | Black | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | % testees selecting each option | | | | | p | % testees selecting each option | | | | | p | % testees selecting each option | | | | | p |
| | 1 | 2 | 3 | 4 | 5 | | 1 | 2 | 3 | 4 | 5 | | 1 | 2 | 3 | 4 | 5 | |
| 46 | 6,3 | 7,7 | 24,7 | 20,9 | 40,4* | 0,40 | 8,1 | 8,3 | 23,4 | 28,4 | 31,9* | 0,32 | 10,3 | 11,0 | 36,2 | 19,4 | 23,1* | 0,23 |
| 47 | 30,5* | 12,7 | 27,3 | 20,6 | 8,9 | 0,31 | 22,8* | 10,8 | 34,7 | 19,4 | 12,3 | 0,23 | 16,7* | 10,5 | 46,2 | 18,0 | 8,7 | 0,17 |
| 48 | 2,4 | 2,2 | 1,7 | 90,9* | 2,8 | 0,91 | 4,0 | 4,2 | 4,6 | 81,4* | 5,7 | 0,81 | 16,7 | 11,2 | 9,6 | 54,9* | 7,7 | 0,55 |
| 49 | 6,7 | 5,8 | 20,2 | 25,6 | 41,7* | 0,42 | 9,5 | 12,6 | 20,0 | 22,1 | 35,7* | 0,36 | 13,1 | 23,8 | 22,0 | 18,1 | 23,1* | 0,23 |
| 50 | 12,0 | 10,8 | 8,0 | 61,5* | 7,7 | 0,62 | 14,1 | 11,8 | 12,5 | 50,8* | 10,7 | 0,51 | 16,7 | 19,4 | 21,2 | 27,4* | 15,4 | 0,27 |
| 51 | 77,4* | 0,5 | 8,4 | 1,6 | 12,0 | 0,77 | 51,0* | 7,2 | 20,3 | 4,6 | 16,9 | 0,51 | 11,3* | 38,0 | 27,4 | 14,0 | 9,4 | 0,11 |
| 52 | 12,1 | 18,9 | 20,3 | 13,7 | 35,0* | 0,35 | 24,1 | 15,2 | 15,3 | 9,9 | 35,5* | 0,36 | 18,0 | 31,6 | 20,0 | 11,6 | 18,9* | 0,19 |
| 53 | 52,7* | 14,4 | 4,9 | 17,2 | 10,8 | 0,53 | 40,8* | 17,3 | 10,2 | 19,3 | 12,4 | 0,41 | 23,4* | 21,1 | 13,2 | 32,4 | 9,9 | 0,23 |
| 54 | 51,6* | 27,1 | 6,1 | 5,8 | 9,4 | 0,52 | 40,0* | 21,7 | 11,6 | 5,9 | 20,9 | 0,40 | 25,5* | 23,8 | 12,3 | 10,0 | 28,4 | 0,26 |
| 55 | 10,8 | 3,0 | 14,9 | 12,1 | 59,2* | 0,59 | 14,0 | 3,3 | 24,2 | 14,1 | 44,4* | 0,44 | 10,4 | 8,1 | 31,1 | 35,0 | 15,4* | 0,15 |
| 56 | 79,2* | 7,1 | 1,3 | 1,0 | 11,3 | 0,79 | 70,8* | 13,6 | 2,4 | 1,0 | 12,2 | 0,71 | 41,8* | 18,0 | 7,4 | 7,6 | 25,3 | 0,42 |
| 57 | 29,8* | 4,7 | 48,2 | 5,7 | 11,6 | 0,30 | 27,9* | 5,5 | 45,7 | 6,5 | 14,3 | 0,28 | 14,1* | 15,2 | 35,5 | 10,3 | 24,8 | 0,14 |
| 58 | 40,1* | 4,9 | 8,8 | 27,9 | 18,2 | 0,40 | 29,2* | 7,7 | 14,2 | 24,4 | 24,5 | 0,29 | 10,7* | 11,9 | 32,3 | 19,9 | 25,1 | 0,11 |
| 59 | 4,4 | 24,9 | 47,9* | 6,9 | 16,0 | 0,48 | 6,4 | 30,2 | 29,6* | 12,2 | 21,6 | 0,30 | 13,1 | 24,1 | 30,4* | 19,4 | 13,0 | 0,30 |
| 60 | 11,4 | 59,6* | 1,4 | 3,1 | 24,5 | 0,60 | 13,2 | 44,7* | 2,8 | 9,4 | 29,9 | 0,45 | 11,1 | 9,5* | 22,9 | 39,2 | 17,4 | 0,10 |
| 61 | 15,7 | 8,2 | 10,1 | 50,6* | 15,6 | 0,51 | 15,9 | 13,2 | 20,2 | 31,2* | 19,4 | 0,31 | 19,0 | 27,1 | 26,6 | 13,3* | 13,9 | 0,13 |
| 62 | 36,0 | 18,9 | 25,2* | 13,1 | 6,8 | 0,25 | 35,4 | 24,3 | 15,6* | 19,7 | 5,0 | 0,16 | 26,5 | 29,3 | 13,8* | 26,0 | 4,5 | 0,14 |
| 63 | 11,7 | 17,7 | 13,1 | 39,2* | 18,2 | 0,39 | 14,1 | 18,9 | 16,9 | 22,5* | 27,6 | 0,23 | 15,4 | 18,1 | 27,5 | 19,4* | 19,5 | 0,19 |
| 64 | 32,5* | 11,7 | 21,6 | 8,5 | 25,7 | 0,33 | 23,6* | 17,5 | 19,4 | 11,3 | 28,2 | 0,24 | 12,5* | 20,9 | 23,8 | 11,9 | 30,9 | 0,13 |
| 65 | 26,0* | 5,3 | 49,6 | 7,1 | 12,1 | 0,26 | 26,2* | 6,8 | 46,4 | 6,8 | 13,8 | 0,26 | 13,6* | 15,4 | 40,1 | 10,9 | 20,0 | 0,14 |
| 66 | 31,2* | 22,7 | 21,6 | 4,3 | 20,2 | 0,31 | 19,7* | 16,0 | 19,6 | 13,7 | 31,0 | 0,20 | 8,2* | 13,1 | 16,4 | 33,5 | 28,8 | 0,08 |
| 67 | 15,9 | 21,0 | 39,6* | 11,4 | 12,1 | 0,40 | 17,7 | 19,3 | 31,4* | 15,4 | 16,2 | 0,31 | 23,9 | 20,2 | 25,0* | 15,8 | 15,1 | 0,25 |
| 68 | 22,9 | 6,9 | 8,7 | 49,4* | 12,1 | 0,49 | 23,8 | 12,3 | 10,1 | 38,3* | 15,6 | 0,38 | 26,6 | 25,4 | 14,7 | 19,1* | 14,2 | 0,19 |
| 69 | 15,3 | 12,9 | 37,7* | 6,8 | 27,3 | 0,38 | 21,4 | 19,6 | 23,3* | 11,8 | 23,8 | 0,23 | 19,8 | 30,6 | 13,8* | 20,5 | 15,4 | 0,14 |
| 70 | 21,1 | 19,3 | 13,2 | 33,1* | 13,3 | 0,33 | 27,1 | 21,9 | 13,7 | 24,7* | 12,6 | 0,25 | 23,2 | 23,4 | 26,3 | 15,7* | 11,4 | 0,16 |
| 71 | 14,1 | 39,6* | 18,6 | 8,7 | 19,1 | 0,40 | 17,2 | 28,0* | 21,3 | 9,0 | 24,4 | 0,28 | 11,2 | 16,4* | 29,9 | 19,8 | 22,7 | 0,16 |
| 72 | 14,9 | 13,4 | 56,6* | 6,7 | 8,4 | 0,57 | 12,2 | 12,0 | 51,6* | 11,2 | 13,0 | 0,52 | 15,6 | 13,9 | 24,0* | 21,1 | 25,5 | 0,24 |
| 73 | 6,2 | 67,6* | 9,0 | 4,5 | 12,7 | 0,68 | 6,8 | 50,9* | 10,3 | 5,0 | 26,9 | 0,51 | 14,6 | 22,1* | 20,7 | 10,5 | 32,1 | 0,22 |
| 74 | 20,6 | 64,0* | 3,9 | 8,8 | 2,7 | 0,64 | 16,7 | 50,5* | 6,1 | 24,1 | 2,6 | 0,51 | 19,0 | 22,6* | 12,5 | 30,5 | 15,3 | 0,23 |
| 75 | 7,3 | 4,1 | 9,4 | 13,4 | 65,8* | 0,66 | 11,0 | 7,5 | 35,3 | 9,1 | 37,1* | 0,37 | 11,0 | 26,2 | 40,1 | 12,0 | 10,7* | 0,11 |
| 76 | 16,6 | 25,3 | 37,5* | 11,6 | 8,9 | 0,38 | 27,4 | 28,4 | 16,3* | 11,6 | 16,3 | 0,16 | 46,4 | 15,1 | 12,8* | 8,1 | 17,6 | 0,13 |
| 77 | 12,1 | 8,7 | 62,3* | 10,8 | 6,1 | 0,62 | 14,0 | 10,0 | 55,2* | 12,7 | 8,1 | 0,55 | 18,5 | 11,3 | 34,7* | 20,4 | 15,1 | 0,35 |
| 78 | 32,2 | 48,8* | 5,7 | 2,2 | 11,1 | 0,49 | 33,0 | 29,1* | 9,7 | 3,7 | 24,4 | 0,29 | 13,8 | 23,1* | 16,2 | 14,7 | 32,3 | 0,23 |
| 79 | 29,6 | 52,0* | 9,4 | 5,0 | 4,0 | 0,52 | 29,9 | 38,5* | 13,9 | 7,4 | 10,2 | 0,39 | 24,3 | 21,1* | 23,9 | 12,5 | 18,2 | 0,21 |
| 80 | 19,0 | 31,8 | 23,9* | 3,5 | 21,8 | 0,24 | 24,2 | 39,1 | 13,0* | 5,8 | 17,9 | 0,13 | 51,4 | 13,2 | 7,4* | 4,0 | 24,0 | 0,07 |
| 81 | 11,1 | 21,7 | 23,8 | 6,3 | 37,0* | 0,37 | 8,0 | 24,7 | 19,7 | 11,2 | 36,4* | 0,36 | 13,8 | 30,3 | 23,5 | 10,3 | 22,0* | 0,22 |
| 82 | 2,3 | 13,4 | 7,7 | 5,9 | 70,8* | 0,71 | 1,8 | 12,0 | 8,9 | 17,9 | 59,4* | 0,59 | 4,5 | 17,3 | 26,7 | 25,0 | 26,5* | 0,27 |
| 83 | 18,0 | 22,0 | 18,9 | 7,8 | 33,2* | 0,33 | 20,5 | 20,4 | 19,4 | 13,2 | 26,5* | 0,27 | 27,3 | 17,1 | 22,1 | 20,2 | 13,2* | 0,13 |
| 84 | 14,3 | 44,3* | 10,5 | 26,6 | 4,2 | 0,44 | 20,7 | 26,7* | 14,4 | 31,3 | 6,8 | 0,27 | 22,3 | 28,1* | 21,8 | 15,5 | 12,3 | 0,28 |
| 85 | 52,5* | 5,4 | 4,9 | 5,7 | 31,5 | 0,53 | 45,7* | 6,4 | 8,2 | 5,5 | 34,1 | 0,46 | 12,6* | 14,5 | 15,4 | 9,3 | 48,2 | 0,13 |

Note: *Correct answer